# Interactive Journal of Medical Research

Impact Factor (2022): 2.0
Volume 15 (2026) ISSN 1929-073X

## Contents

### Corrigenda and Addendas

Correction: Digital Literacy and Interpersonal Trust as Predictors of Willingness to Share Patient-Generated Health Data Among Korean Internet Users: Cross-Sectional Study Using Privacy Calculus and Communication Privacy Management Theories (e103310)
Dongsu Lee, Wonseuk Jang. . . . 8

Correction: Periorbital Necrotizing Fasciitis: Case Presentation (e108273)
JMIR Editorial Office. . . . 10

Correction: Validation and Refinement of the Sense of Coherence Scale for a French Population: Observational Study (e110805)
Paul Sebo, Benoit Tudrej, Augustin Bernard, Bruno Delaunay, Alexandra Dupuy, Claire Malavergne, Hubert Maisonneuve. . . . 12

### Viewpoints

Generosity as a Scientific Method: Building Knowledge and Community in a Competitive World (e84933)
Bruno Andrade. . . . 14

Rethinking Basic Science Teaching in Medical Education: A Viewpoint on Its Past, Present, and Future (e80646)
Kátia Avena, Ana Brito, Luiz Quintanilha, Mariana Araújo-Pereira, Bruno Andrade. . . . 21

Health Care Analytics Challenges: A 3-Pillar Framework Connecting Analytics Maturity, Workforce Agility, and Technical Enablement (e96541)
Samuel Harrold. . . . 30

### ViewPoint

Evidence-Based Self-Management Strategies for Fibromyalgia: Foundations for Digital Therapeutic Applications (e67523)
Thomas Lebeau Foustoukos, Isabelle Leclercq, Marc Blanchard, Thomas Hügle. . . . 43



XSL•FO
RenderX

## Reviews


How Many Instruments Do We Need? Strengths, Weaknesses, and Potentials for Further Developments of Digital Health Literacy Measurement Instruments: Narrative Overview of Reviews and Content Analysis (e91603)

Pauline Kaboth, Lorenz Harst, Jochen Schmitt, Madlen Scheibe. . . . . 57

Landscape, Evidence, Gaps, and Opportunities in Digital Mental Health Interventions for Older Adults: Scoping Review (e92542)

Leyi Zhou, Xinyi Zuo, Joonyoung Cho, Chuxuan Zheng, Shengyi Jing, Xiaoling Xiang. . . . . 81

Trust Barriers and Vulnerabilities in Older Adults' Telemedicine Adoption: Scoping Review (e84818)

Wanqing Wang, Xin Ma, Matthew Liu, Angela Chang. . . . . 97

Digital Health Interventions Incorporating Behavior Change Techniques and Gamification for Bladder Health in Adults Aged 50 Years and Older: Rapid Review (e76867)

Oscar Aguila-Gimeno, Norina Gasteiger, Emma Stanmore, Marius Brazaitis, Rima Solianik, Erika Karkauskiene, Laura Jarutien , Júlia Romeu-Busquets, Javier Jerez-Roig. . . . . 113

Health Information Technologies in the Support System of New Fathers' Postnatal Depression: Scoping Review (e95238)

Weijie Wang, Erika Penney, Jaime Garcia. . . . . 133

Contexts and Mechanisms Related to the Efficacy of Digital Medication Adherence Interventions to Support Medication Adherence Among Adults With Chronic Diseases: Realist Review (e89803)

Srinjaya Saha, Yvonne Latham, Carol Holland. . . . . 146

Enhancing Healthy Longevity: Scoping Review of Practices and Interventions (e94085)

Angela Leung, Karen Cheung, Kai-ling Ou, Ivy Zhao, Min Qiu, Ka Wan, Jed Montayre, Yaqi Huang. . . . . 172

Factors Influencing Adherence to Digital Lifestyle Interventions for Adolescents: Systematic Review and Meta-Analysis of Attrition (e84822)

Nikolaos Boumparis, Olivia Studhalter, Philippe de Riedmatten, Inci Yücel, Kleio Koutra, Katrina Champion, Ana Molina-Barceló, Teresa de Pablo-Pardo, Haridimos Kondylakis, Michael Schaub, Andreas Triantafyllidis, Severin Haug. . . . . 193

Design, Facilitation, and Evaluation of Tabletop Exercises for Prehospital Mass Casualty Preparedness: Scoping Review (e96228)

Paurnami Prashanth, Ali AlRahma, Dayol Narayanan, Abu Omayer, Azza Yousif, Ives Hubloue, Nabil Zary. . . . . 217

Application of Machine Learning Algorithms to Health Systems Planning and Public Policy Using Real-World Data: Scoping Review (e96779)

Deusa Rocha, Gentil Barbosa, Paulo De Souza Bermejo. . . . . 236

Impact of Elevated Blood Pressure Targets on Splanchnic Ischemia and Mortality in Septic Resuscitation: Systematic Review (e98772)

Paulina Cisneros, John Ortiz Caviedes, Lina Piñeros Ladino, Natalia Quintero, Valeria Oñate Cervera, Juan Urrea Ospina, Erika Cerna Alcantar.
2 4 8

Overcrowding Indicators in Emergency Departments Across Countries: Scoping Review (e78073)

Natasya Mohd Nasir, Ku Indera Putera, Nur Noris, Zalina Libasin, Muniamal Krishnan, Nor Salaton, Kah Lum, Nur Abdullah, Intan Saimy. . .
2 6 4





XSL•FO
RenderX

Information and Communication Technologies for Chronic Disease Self-Management in Adults Aged 65 Years and Older: Scoping Review (e60542)
Paul Murdock, Yiyi Wu, Charles Senteio. . . . 281

Global Research Trends and Hotspots in Gene Editing and Stem Cell Therapies for Neurodegenerative Diseases: Bibliometric and Visualization Analysis (e83709)
Lijun Xiang, Yun Xiao, Ming Cai, Jing Qin, Ting Wang, Xueming Xiang, Jun Ke, Ganlin Peng. . . . 295

Mapping Digital Nudges and Recommender Systems for Obesity Prevention: Scoping Review (e73151)
Sarah Forberger, Lucia Reisch, Pieter Van Gorp, Christoph Stahl, Lara Christianson, Jihan Halimi, Karina De Santis, Chungwan Lo, Cassandra Omane, Alejandra Loyola-Leyva, Laurent Malisoux, Tiziana de-Magistris, Torsten Bohn. . . . 317

Emerging Technologies and Vulnerabilities in Older Adults Without Cognitive Impairments: Systematic Review of Qualitative Evidence (e69676)
Annachiara Fasoli, Maria De Luca, Giorgia Beretta, Chris Gastmans, Virginia Sanchini. . . . 340

Exploring Serious Games in Supporting Postnatal Depression: Narrative Review (e70777)
Weijie Wang, Erika Penney, Valerie Gay, Jaime Garcia. . . . 363

Identifying Key Predictors of Appropriate Discharge Destinations for Older Inpatients in Acute Care: Scoping Review (e76582)
Christoph Leinert, Marina Fotteler, Thomas Kocar, Jessica Wolf, Lisa Beissel, Kathrin Grummich, Dhayana Dallmeier, Michael Denkinger. . . .
3 8 6

The Effectiveness, Facilitators, and Barriers of Digital Mental Health Services for First Nations People in Australia: Systematic Scoping Review (e80386)
Siyu Zhai, Andrew Goodman, Anthony Smith, Sandra Diminic, Xiaoyun Zhou. . . . 405

The Usability of Continuous Monitoring Devices With Deterioration Alerting Systems in Noncritical Care Units: Scoping Review (e75713)
Jo-Fan Pan, Dawn Dowding, David Wong, Ashley Scott, Qimeng Zhao. . . . 438

Machine and Deep Learning for Detection of Moderate-to-Vigorous Physical Activity From Accelerometer Data: Systematic Scoping Review (e76601)
Yahua Zi, Sjors van de Ven, Eco de Geus, Peijie Chen. . . . 456

Treatment of Gender in Research on Intervention Programs Targeting Social Isolation and Loneliness Among Older Adults: Scoping Review (e72281)
Kenta Nomura, Naoto Kiguchi, Eisuke Inomata, Takeshi Nakamachi, Norikazu Kobayashi. . . . 482

Internet Use and Self-Rated Health Among Older Adults: Scoping Review (e76930)
Hanan AboJabel, Fareeda Abo-Rass. . . . 497

Treating Behavioral Addictions With Augmented Reality and Virtual Reality: Scoping Review (e77011)
Felicia Chiew, Genevieve Li, Victrine Tseung, Jing Shi. . . . 518

Constructs Influencing Patient Perceptions of Use of AI in Medical Imaging Analysis: Systematic Review (e92969)
Preksha Kuppanda, Monika Janda, Liam Caffery. . . . 531

Digital Interventions Addressing Cognitive and Psychological Symptoms in Long COVID: Scoping Review of Multicomponent Approaches (e80616)
Sandra León-Herrera, Marta Sánchez-Castro, Ana Luisa Neves, Mª Rodríguez-Pérez, Vinicius Jobim Fischer, Djenna Hutmacher, Reham Aldakhil, Marina Vaillancourt de Dios, Vinicius Anjos de Almeida, Bárbara Oliván-Blázquez, Rosa Magallón-Botaya, Charles Benoy, Raquel Gómez-Bravo. . . . 578



XSL•FO
RenderX

## Tutorials

An Introduction to AI for Clinicians: Tutorial (e85266)
Stephen Lee, Alexis Carter, Muhammad Haider, Seok-Bum Ko. . . . 545

Enhancing Transparency, Auditability, and Reproducibility of Deduplication in Systematic Reviews: Tutorial for the Rayyan Method and Systematic Auto Resolver (e94317)
Mimi Kim, Amin Shanaa, Hossam Hammady, Mohammad Aboelnour, Robert Ayan. . . . 555

## Original Papers

Real-World Clinical Characterization of Major Depressive Disorder and Treatment-Resistant Depression Supported by Natural Language Processing: Multicenter Observational Study From the MOOD Project (e86448)
Dieter Zeeuws, Katrien Bernagie, Geert De Bruecker, Isabel Claeys, Charlotte Evenepoel, Fabienne Ver Donck, David Smeets, Elke Peeters.
5 6 4

Pulse Discovery Toolkit, a Multicomponent Nutrition Intervention for Preschool Children in Childcare Centers: Mixed Methods Pilot Study (e55866)
Hiwot Haileslassie, Renee Ramikie, Hassan Vatanparast, D Ramdath, Amanda Froehlich Chow, Phyllis Shand, Rachel Engler-Stringer, Jessica Lieffers, Shannon Hood-Niefer, Carol Henry. . . . 591

Prevalence and Associated Factors of Excessive Dietary Supplement Use Among Japanese Adults: Cross-Sectional Study (e82623)
Minami Sugimoto, Keiko Asakura, Nana Shinozaki, Kentaro Murakami. . . . 606

Determining the Relationship Between People's Explicit and Implicit Preferences for Gender-Inclusive Sexual and Reproductive Health Content: Randomized Controlled Trial (e85868)
Elizabeth Boskey, Jessica Kant, Ariel Berman, Frances Grimstad. . . . 625

Association Between Depressive Symptoms and Incidence of Stroke in a Population With Cardiovascular-Kidney-Metabolic Syndrome Stages 0 to 3: Nationwide Prospective Cohort Study (e80988)
Yunjie Li, Lin Wang, Lujing Gao. . . . 640

Patient-Reported Quality of Life in Lung Transplantation: Prospective Cohort Study (e86211)
Peng Li, Jiawei Wang, Hongyi Wang, Shuo Li, Lei Wang, Shan Gao, Chunxiao Hu, Guangjian Zhang. . . . 656

Cardiac Resynchronization Therapy With Defibrillator Using the JROAD-DPC Database: Cost-Effectiveness Analysis (e94073)
Kazuki Ohashi, Masaya Watanabe, Yasuhiro Morii, Hisashi Yokoshiki, Kengo Kusano, Katsuhiko Imai, Masahiko Takagi, Teiichi Yamane, Hiroshi Tada, Katsuhiko Ogasawara. . . . 668

Comparative Analysis of Cardiometabolic Risk Profiles Among Hypertensive and Normotensive Adults: Cross-Sectional Study (e84466)
Lebogang Thaga, Gudani Mukoma, Terry Ellapen, Takalani Muluvhu. . . . 680

The Lifetime Dance Exposure Questionnaire for Professional Training: Survey Development and Reliability Study (e75632)
Aston McCullough, Kai Han. . . . 692



XSL•FO
RenderX

Understanding the Functional Needs of Patients With Cardiovascular Disease Regarding Nursing Robots Using a Kano Survey Scale: Cross-Sectional Study (e92610)
XiuLi Wu, Aimei Kang. . . . 706

Identification of the Core Competencies Required in Endodontics for Undergraduate Students in Syrian Dental Schools by Using a Modified Delphi Technique: Prospective Exploratory Survey Study (e83799)
Muhammad Salameh, Mayssoon Dashash, Issam Jamous. . . . 718

COVID-19 Rebound in Nirmatrelvir Plus Ritonavir Treatment and Control Groups: Prospective Cohort Study (e80263)
Jacqueline Kueper, Kalyani Kottilil, Giorgio Quer, Danielle Chiang, Emily Spencer, Jyothi Purushotham, Edward Ramos, Leila Roumani, Kristian Andersen, Eric Topol, Jay Pandit, Michael Mina. . . . 727

Factors Affecting the Accuracy of Clinical Coding in the Casemix System in a Teaching Hospital: Cross-Sectional Study (e93769)
Amirah Azzeri, Wan Wan Musaludin, Hafiz Jaafar, Nurnabihah Hafidz, Azimatun Noor Aizuddin. . . . 739

Individual Research Behaviors and Research Funding Acquisition Across Fields and Career Periods: Regression Analysis (e98428)
Akiko Hashiguchi, Satoru Takahashi. . . . 749

Innovation Deimplementation in Emergency Departments During the COVID-19 Pandemic: Qualitative Study of Clinicians' Decision-Making (e82088)
Shreya Huilgol, Nabeel Qureshi, Carl Berdahl, Catherine Cohen, Peter Mendel, Shira Fischer. . . . 770

Sociodemographic and Health-Risk Determinants of COVID-19 Vaccine Booster Preferences and Willingness to Pay in Singapore: Discrete Choice Experiment (e87909)
Hooi Cheng, Yi Wang, Sharon Tan, Ian Ang. . . . 783

Health Care–Related Determinants of First-Time Long-Term Care Need in Older Adults in Germany: Retrospective Cohort Study Using Claims Data (e86572)
Till Baldenius, Kathrin Jürchott, Christine Haeger, Susann Behrendt, Stefan Blüher, Susanne Schnitzer. . . . 799

Barriers and Facilitators to Physician-Patient Communication in Chinese Tertiary Hospitals From the Perspectives of Hospital-Based Physicians and Patient Relations Coordinators: Qualitative Study (e87947)
Zengping Shi, Qinqin Jiang, Xincheng Wang, Hongli Yan, Yi Xia, Lushaobo Shi, Dong Wang. . . . 815

Codeveloping a Clinical Pathway for Recurrent Clostridioides difficile Infection With Primary Care Providers: User-Centered Design Study (e93032)
Ritika Patel, Rose Franz, Jordan Tate, Jennifer Coppens, Matthew Do Rego, Christopher Smith, Lauretta Ovadje, Karen Wong, Lynora Saxinger, Dina Kao. . . . 829

Impact of mHealth-Augmented Social Support on Health Care Use Among Patients With Diabetes: Secondary Analysis of the TExT-MED+FANS Trial (e65113)
Danielle Hazime, Liza Raffi, Elizabeth Burner. . . . 840

Impact of a New Virtual Health Humanities Approach for Collaborative Online International Learning: Mixed Methods Study (e84474)
Heather Gaunt, Kwang Cham, Florian Knothe, Chee Lim, Rosie Shea, Hui Chng. . . . 852

Classroom Integration of AI-Simulated Patients and Automated Feedback in Doctoral Psychotherapy Training: Longitudinal Mixed Methods Study (e93199)
Ellen Converse, Ryan Louie, Debra Safer, Juan Pacheco, William Fang, Jamie Kent, Bruce Arnow, Diyi Yang. . . . 870



XSL•FO
RenderX

Repeat At-Home Saliva Collection to Measure Cortisol Awakening Response (CAR): Protocol Feasibility Study Among Black Women With Histories of Abuse (e64150)
Kiyomi Tsuyuki, Marguerite Lucea, Andrea Cimino, Michael Killian, Christina Catabay, Engle Abrams, Emmanuel Baffour-Siaw, Jacquelyn Campbell, Douglas Granger, Jamila Stockman. . . . 896

Effectiveness and Lessons Learned From an Occupational E-Mental Health Intervention for Enhancing Workplace Mental Health: The EMPOWER Cluster Randomized Controlled Trial (e66041)
Carlota de Miquel, Christina Van der Feltz-Cornelis, Leona Hakkaart-van Roijen, Dorota Merecz-Kot, Marjo Sinokki, Jordi Rodeiro-Boliart, Jennifer Sweetman, Kaja Staszewska, Ellen Vorstenbosch, Daniele Porricelli, Stijn Peeters, José Ayuso-Mateos, Luis Salvador-Carulla, Sue Lukersmith, Oriol Borrega, Carla Sabariego, Christophe Vanroelen, Alberto Raggi, Diletta Porcheddu, Josep Haro, Beatriz Olaya. . . . 911

Perspectives of Indian Gastroenterologists and Hepatologists on Nonalcoholic Fatty Liver Disease Diagnosis and Management: Insights From the Nationwide Web-Based Cross-Sectional DRIVE Survey (e75138)
Anil Arora, Yogesh Garje, Shagupta Shaikh, Shruti Dharmadhikari, Chintan Khandhedia, Neeraj Markandeywar, Amey Mane, Suyog Mehta, Ashish Kumar. . . . 930

Trends, Predictors, and Outcomes of Monitored Acute Care Unit Admissions in Older Adults: 10-Year Retrospective Analysis (e80629)
Ludovic van Delden, Stephan von Düring, Frederic Rouyer, Anne Rossel, Christophe Marti, Sara Cereghetti, Xavier Roux, Thibaut Desmettre, Olivier Grosgurin, Christophe Fehlmann. . . . 942

Frailty Screening and Management for Older Australians in General Practice: Mixed Methods Evaluation (e79681)
Jennifer Job, Caroline Nicholson, Ruby Strauss, Debra Clark, Anita Pelecanos, Claire Jackson. . . . 959

Modeling Frailty Risk Based on Biochemical and Hematological Parameters Using Multiple and Logistic Regression: Cross-Sectional Analysis of the Toledo Study for Healthy Aging (e76500)
Francisco Zamudio, Diego Arauna, Francisco Garcia, Cecilia Albala, Victoria Villalobos, Eduardo Fuentes, Ivan Palomo. . . . 976

COPD-SQ Positivity and Suspected Asthma Screen Positivity From a Representative, Community-Based Sample in Xishui County, Guizhou Province, China: Cross-Sectional Study (e108291)
Yuhao Liu, Shiyu Zhang, Zhoutao Zheng, Ke Huang, Qiande Lai, Yingping Wang, Tianying Zhao, Yingchi Luo, Xiangqin Lyu, Pascal Geldsetzer, Till Bärnighausen, Ting Yang, Simiao Chen, Chen Wang. . . . 998

Value of Blood Count–Derived Inflammatory Markers for Evaluating Psoriasis Severity: Pilot Cross-Sectional Observational Study (e86454)
Luyuan Wang, Xiaorui Zhang, Chuangang Xia, Wenyu Li, Qiuju Li, Wen Xiao, Youkun Lin. . . . 1014

Mapping the Global Surge in Postoperative Sleep Research From 2014 to 2024: Bibliometric Analysis (e86086)
Xin Wei, Lijuan Fu, Wencai Jiang, Xianjie Zhang, Rui Zhou. . . . 1026

Sleep Disturbance and Its Association With Purchasing Behavior of COVID-19 Medicine Among the Public After the Adjustment of Zero-COVID Policy in China: Results From a Web-Based Survey Study (e79903)
Yu Huang, Xiang Zhao, Lei Wang, Qiaohong Lv, Qingqing Wu, Shuiyang Xu, Xuehai Zhang, Shuxian Wu. . . . 1038

Clinical Effectiveness and Cost-Effectiveness of Blended Treatment for Major Depression Compared With Treatment as Usual Within Routine Care in Europe: A Noninferiority Randomized Controlled Trial (e72866)
Heleen Riper, Annet Kleiboer, Adriaan Hoogendoorn, Judith Bosmans, Mieke Schulte, Kim Mathiasen, Jean Hazo, Jerome Holtzmann, Karine Chevreul, David Ebert, Ingrid Titzler, Mathias Berking, Burkhardt Funk, Anneke van Schaik, Lise Kemmeren, Roman Cieslak, Ewelina Smoktunowicz, Anna Maj, Cristina Botella, Azucena Garcia-Palacios, Rosa Banos, Rocío Herrero, Gerhard Andersson, Naira Topooco, Kristofer Vernmark, Thomas Berger, Tobias Krieger, Arlinda Cerga-Pashoja, Asmae Doukani, Ricardo Araya, Eirini Karyotaki, Spyros Kolovos, Mark Hoogendoorn, Ward van Breda, Pepijn van de Ven, Artur Rocha, Gonçalo Gonçalves, Pim Cuijpers. . . . 1051



XSL•FO
RenderX

Burden of Insomnia Among Individuals With Major Depressive Disorder: Cross-Sectional Observational Study (e93231)
Josh Hamilton, Emilie Pain, Jesper Riise, Zhiheng Zhang. . . . 1078

Effects of a Personalized Retrieval-Augmented Generation Chatbot on Information Needs in Patients With Chronic Kidney Disease: Mixed Methods Randomized Controlled Trial (e96501)
Hee Hwang, Junghwan Kim, Dongseok Heo, Jeong You, Jeonghwan Lee, Soie Kwon, Kyung Yoo, Jung Yoon Park, Bongwon Suh, Jung Lee, Ki Lee. . . . 1145

Intravenous N-Acetylcysteine in Patients With Dengue Fever With Marked Aminotransferase Elevation: Prospective Comparative Study (e97990)
Huyen Nguyen, Lien Tran, Hieu Nguyen, Trang Phan, The Ngo. . . . 1169

## Case Reports

Patients With Acute Stroke Symptoms Consulting Large Language Model Chatbots Before Presenting to the Emergency Department: Case Series (e107318)
Sara Dierickx, Paolo Speziali, Helena Gossye, Sofia Slootjes, Sylvie De Raedt, Anaïs Missiaen, Michel Van Lint, Edwin Van Dellen, Fenne Vandervorst. . . . 1103

An Eye-Opening Approach: Cancer of Unknown Primary Source With Choroidal Metastasis Case Report (e77895)
Erin Kaser, Vedin Barve, Alexander Blaschke, Constance Fry, Corey Waldman, Sarah Hackman, William Kelly. . . . 1109

Orthodontic Management of Vertically Impacted Mandibular Permanent Second Molar in Adults: Case Report (e101851)
Mashaka Mrisho, Cui Lingling, Mary Pallangyo, Jiaao Li, Deogratius Galikano, Xinwei Guo, Yue Zhang, Si Xiao, Yan Zhang. . . . 1118

## Research Letters

The Shifting Landscape of Natural Health Product Licensing in Canada: The e-Commerce Case of Stem Cell Supplements (e86747)
Alessandro Marcon, Michael Jewer, Timothy Caulfield. . . . 1129

Sustained Health Care–Associated Infection Reduction Among Electronic Hand Hygiene Monitoring System Clients (e101905)
Gabrielle Genovesi. . . . 1134

Within-Provider Variation in Prices for Commonly Utilized Services in Office Setting: Cross-Sectional Study (e96883)
Yuvraj Pathak, David Muhlestein. . . . 1138

## Retraction

Retraction: Antibiotic Prescribing Behavior of Physicians in Outpatient Departments in Hospitals in Northwest Ethiopia: Structural Equation Modeling Approach (e105480)
JMIR Editorial Office. . . . 1178



XSL•FO
RenderX

# Correction: Digital Literacy and Interpersonal Trust as Predictors of Willingness to Share Patient-Generated Health Data Among Korean Internet Users: Cross-Sectional Study Using Privacy Calculus and Communication Privacy Management Theories

Dongsu Lee[1], MPH; Wonseuk Jang[2], PhD

[1]Future Medicine Research Center, Gangnam Severance Hospital, College of Medicine, Yonsei University, Seoul, Republic of Korea

[2]Department of Medical Device Engineering and Management, Yonsei University College of Medicine, 50-1, Yonsei-Ro, Seodaemun-gu, Seoul, Republic of Korea

**Corresponding Author:**
Wonseuk Jang, PhD
Department of Medical Device Engineering and Management, Yonsei University College of Medicine, 50-1, Yonsei-Ro, Seodaemun-gu, Seoul, Republic of Korea

**Related Article:**

Correction of: https://www.jmir.org/2026/1/e75448

***(Interact J Med Res 2026;15:e103310)*** doi:10.2196/103310

In "Digital Literacy and Interpersonal Trust as Predictors of Willingness to Share Patient-Generated Health Data Among Korean Internet Users: Cross-Sectional Study Using Privacy Calculus and Communication Privacy Management Theories" [1], the authors noted one error.

The affiliation of the corresponding author has been corrected. The following affiliation was revised:

*Future Medicine Research Center, Gangnam Severance Hospital*

*College of Medicine*

*Yonsei University*

*20 63rd Road*

*Unju-ro, Gangnam-gu*

*Seoul, 06229*

*Republic of Korea*

The new information was added as affiliation 2, and reads:

*Department of Medical Device Engineering and Management*

*Yonsei University College of Medicine*

*50-1, Yonsei-Ro, Seodaemun-gu*

*Seoul, 03722*

*Republic of Korea*

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

## Reference

1. Lee D, Jang W. Digital literacy and interpersonal trust as predictors of willingness to share patient-generated health data among Korean internet users: cross-sectional study using privacy calculus and communication privacy management theories. J Med Internet Res 2026 May 27;28:e75448. [doi: 10.2196/75448] [Medline: 42204765]

*Submitted 02.Jun.2026; this is a non–peer-reviewed article;accepted 10.Jun.2026; published 31.Jul.2026.*

*Please cite as:*
*Lee D, Jang W*
*Correction: Digital Literacy and Interpersonal Trust as Predictors of Willingness to Share Patient-Generated Health Data Among Korean Internet Users: Cross-Sectional Study Using Privacy Calculus and Communication Privacy Management Theories*
*Interact J Med Res 2026;15:e103310*
*URL: https://www.i-jmr.org/2026/1/e103310*
*doi:10.2196/103310*



XSL•FO
RenderX

© Dongsu Lee, Wonseuk Jang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 31.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Corrigenda and Addenda

# Correction: Periorbital Necrotizing Fasciitis: Case Presentation

JMIR Publications Editorial Office[1]

JMIR Publications, Toronto, ON, Canada

**Corresponding Author:**
JMIR Publications Editorial Office
JMIR Publications
130 Queens Quay E, Suite 1100
Toronto, ON M5A 0P6
Canada
Phone: 1 416 583 2040
Email: ed-support@jmir.org

**Related Article:**

Correction of: https://www.i-jmr.org/2023/1/e52507

***(Interact J Med Res 2026;15:e108273)*** doi:10.2196/108273

Following the publication of "Periorbital Necrotizing Fasciitis: Case Presentation" [1], concerns were raised regarding a potential conflict of interest between reviewer Karen Wong Riff and the authors. This concern included shared affiliation with members of the author list. This conflict was not disclosed to the JMIR Publications Editorial Office as per journal policy.

This article has been subsequently reassessed by the editor in chief, who determined that the content of this review did not unduly influence the decision to accept the article. The decision to accept the submission was based primarily on the remaining reviewers' comments and the editor in chief's contemporary appraisal of the manuscript.

As such, the JMIR Publications Editorial Office issues this corrigendum to update the Conflicts of Interest statement:

> *Reviewer Karen Wong Riff was found to have an undisclosed conflict of interest with members of the author list under the terms of JMIR Publications policy. The editor in chief determined that the content of the review did not unduly influence the decision to accept the article. The decision to accept the submission was based primarily on the remaining reviewers' comments and the editor in chief's contemporary appraisal of the manuscript.*

The JMIR Publications Editorial Office further issues an additional corrigendum to update the informed consent information, which is included in the "Ethical Considerations" subsection as follows:

*Informed consent was obtained, in writing, from the patient for this publication.*

We regret that these issues were not identified before publication.

The correction will appear in the online version of the paper on the JMIR Publications website together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

## Reference

1. Huang RS, Patil NS, Khan Y. Periorbital necrotizing fasciitis: case presentation. Interact J Med Res 2023 Nov 28;12:e52507 [FREE Full text] [doi: 10.2196/52507] [Medline: 37971729]



XSL•FO
RenderX

*Submitted 29.Jul.2026; this is a non–peer-reviewed article;accepted 03.Aug.2026; published 18.Aug.2026.*

*Please cite as:*
*Editorial Office JMIRP*
*Correction: Periorbital Necrotizing Fasciitis: Case Presentation*
*Interact J Med Res 2026;15:e108273*
*URL: https://www.i-jmr.org/2026/1/e108273*
*doi:10.2196/108273*
*PMID:42612217*

©JMIR Publications Editorial Office. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 18.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Corrigenda and Addenda

# Correction: Validation and Refinement of the Sense of Coherence Scale for a French Population: Observational Study

Paul Sebo[1], MSc, MD; Benoit Tudrej[2], MD, PhD; Augustin Bernard[1], MD; Bruno Delaunay[1], MD; Alexandra Dupuy[1], MD; Claire Malavergne[1], MD; Hubert Maisonneuve[1,2], MD

[1]Institut universitaire de Médecine de Famille et de l'Enfance (University Institute for Primary Care), University of Geneva, Geneva, Switzerland
[2]University College of General Medicine, University Claude Bernard Lyon 1, Lyon, France

**Corresponding Author:**
Paul Sebo, MSc, MD
Institut universitaire de Médecine de Famille et de l'Enfance (University Institute for Primary Care)
University of Geneva
Rue Michel Servet 1
Geneva, 1211
Switzerland
Phone: 41 22 379 50 61
Email: paulsebo@hotmail.com

**Related Article:**

Correction of: https://www.i-jmr.org/2024/1/e50284

***(Interact J Med Res 2026;15:e110805)*** doi:10.2196/110805

In “Validation and Refinement of the Sense of Coherence Scale for a French Population: Observational Study" [1], information regarding the permission to use the Sense of Coherence (SOC) scales and their copyright was not included.

In the Validation of the French Version of the 3-Dimension SOC-13 and Development of the Unidimensional 8-Item SOC Questionnaire section of the Methods, the following paragraph was revised:

> *The French versions of the SOC-13 and the new unidimensional 8-item SOC scale (SOC-8) are provided in Multimedia Appendix 1. Following published guidelines, we targeted a minimum of 500 participants, achieving a “very good” sample size, with a responder-to-item ratio exceeding 20:1 [12]. All analyses were performed with Stata 15.1. [Multimedia Appendix 1]*

The paragraph now reads:

> *The French versions of the SOC-13 and the new unidimensional 8-item SOC scale (SOC-8) are provided in Multimedia Appendix 1. The SOC questionnaire was used with permission from the copyright holder via the STARS (Society for Theory and Research on Salutogenesis) website. Permission to use the SOC scales must be obtained through STARS. Following published guidelines, we targeted a minimum of 500 participants, achieving a “very good” sample size, with a responder-to-item ratio exceeding 20:1 [12]. All analyses were performed with Stata 15.1.*

The caption for Multimedia Appendix 1 was revised from the following:

> *The French versions of the SOC-13 and the SOC-8.*

The caption now reads:

> *The French versions of the SOC-13 and the SOC-8. Use of these questionnaires without prior permission via the STARS website is prohibited.*

In addition, a visible watermark has been added to the Multimedia Appendix 1 file stating:

> *Copyrighted material, reproduction prohibited.*

The content of the SOC-13 and SOC-8 questionnaires themselves has not been changed. These changes do not affect the results or conclusions of the article.

The correction will appear in the online version of the paper on the JMIR Publications website together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

## Reference



XSL•FO
RenderX

1. Sebo P, Tudrej B, Bernard A, Delaunay B, Dupuy A, Malavergne C, et al. Validation and refinement of the Sense of Coherence Scale for a French population: observational study. Interact J Med Res 2024 Jul 16;13:e50284 [FREE Full text] [doi: 10.2196/50284] [Medline: 39012689]

*Submitted 30.Aug.2026; this is a non–peer-reviewed article;accepted 09.Sep.2026; published 20.Nov.2026.*

*Please cite as:*
*Sebo P, Tudrej B, Bernard A, Delaunay B, Dupuy A, Malavergne C, Maisonneuve H*
*Correction: Validation and Refinement of the Sense of Coherence Scale for a French Population: Observational Study*
*Interact J Med Res 2026;15:e110805*
*URL: https://www.i-jmr.org/2026/1/e110805*
*doi:10.2196/110805*
*PMID:*

©Paul Sebo, Benoit Tudrej, Augustin Bernard, Bruno Delaunay, Alexandra Dupuy, Claire Malavergne, Hubert Maisonneuve. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 20.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Generosity as a Scientific Method: Building Knowledge and Community in a Competitive World

Bruno B Andrade[1,2,3], MD, PhD

[1]Division of Infectious Diseases, Department of Medicine, Johns Hopkins University School of Medicine, 1830 East Monument Street, Room 441, Baltimore, MD, United States

[2]Department of International Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[3]Laboratório de Pesquisa Clínica e Translacional, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil

**Corresponding Author:**
Bruno B Andrade, MD, PhD
Division of Infectious Diseases, Department of Medicine, Johns Hopkins University School of Medicine, 1830 East Monument Street, Room 441, Baltimore, MD, United States

## Abstract

Generosity can function as a scientific method—a disciplined stance that aligns curiosity with openness, credit-sharing, and stewardship of data, specimens, and ideas. Rather than a soft add-on, generosity structures how questions are framed, teams are built, and results are disseminated, thereby improving rigor, reproducibility, and impact. This viewpoint article advances a conceptual and operational framework for "generosity in science," aimed at researchers, institutions, and funders seeking alternatives to competition-driven models of knowledge production. I examine generosity as practice at the levels of people, collectives, and institutions and argue that persistent global challenges in health demand pro-collaborative architectures. Seen this way, generosity is not mere altruism; it is part of the epistemic engine that turns uncertainty into shared knowledge while distributing opportunity and recognition more fairly. I define core principles of generous research and organize them across three domains: research design, governance, and evaluation. The paper draws on illustrative examples and relevant literature to situate generosity within ongoing debates on open science, team science, and research assessment reform. I outline practical principles for embedding generosity into research design, governance, and evaluation and discuss how these principles can counter vanity metrics and short-term incentives. I conclude that embedding generosity in the infrastructure of science enables better questions, faster learning, and greater public value.

***(Interact J Med Res 2026;15:e84933)*** doi:10.2196/84933

**KEYWORDS**

generosity; open science; team science; leadership; mentorship

## Introduction: A Different Currency of Success

Science has always been sold to us as a race. A race against time, disease, ignorance, and, above all, against other scientists. The metaphors are warlike: we "fight" cancer, we "combat" pandemics, we "outpace" competitors. This narrative has been useful for funding agencies, for politicians, and for institutions that thrive on the drama of competition. Yet, beneath this spectacle, the real progress of science often emerges from something far quieter and less glamorous: the capacity of human beings to share, to collaborate, and to give.

The book *Infectious Generosity* by Chris Anderson argues that generosity is a kind of contagion, spreading rapidly once individuals and institutions embrace it as a central value [1]. It has made me reflect on how my own scientific journey was shaped less by the victories of rivalry and more by the power of generosity. My most meaningful contributions have been born from collaborations that began with trust and flourished through the act of giving, whether data, time, mentorship, or simply space for others to shine. But having collaborative environments, colleagues, and agencies willing to share (space, reagents, time, and data) is practically a privilege. A growing body of research has demonstrated that reward and recognition systems in science are not neutral, but stratified by geographic, linguistic, gender, and racial criteria, shaping who is cited, funded, and visible in the global knowledge economy [2-6]. In this context, generosity cannot be understood solely as an individual virtue; it is also constrained and enabled by structural conditions that distribute opportunities unevenly. Thus, fostering generosity requires interpersonal practices within laboratories, but also systemic changes in how scientific contribution and value are recognized.

The central question is unavoidable: what happens if we take generosity seriously as a scientific method? If we treat it as a core value of our work, not as a footnote to CVs or a soft skill for committees? This viewpoint article explores how generosity operates in contemporary science, the conflicts it faces with the



XSL•FO
RenderX

traditional machinery of academic success, how it may contribute to more inclusive, rigorous, and sustainable forms of knowledge production and why future generations must choose genuine generosity (and not just express it in social media posts) if they are to build a better, more sustainable scientific world.

## Generosity in a Scientist's Career

When I look back at my career, the turning points did not come from moments when I "outcompeted" someone, even though an outside observer might initially see it that way. In fact, they came from moments when I decided to share, whether it was reagents, ideas, or time. Early on, I realized that mentorship was not a distraction from science but a way to multiply its reach. I invested time in training students, guiding them through complex questions, and giving them the confidence to pursue independent paths. Some of them are now leading their own groups, writing their own grants, and mentoring the next wave. Their success does not diminish mine; it expands it. Each student mentored is like a small wave meeting the shore: gentle on its own, but when joined with others, they gather momentum until they swell into a tide, a force capable of reshaping coastlines. What begins as individual growth becomes a cascading surge of knowledge, leadership, and impact, where today's student becomes tomorrow's mentor, carrying the motion forward. In this sense, scientific contribution expands through relational networks rather than remaining confined to individual achievement.

A concrete illustration in my career is the Regional Prospective Observational Research in Tuberculosis (RePORT) International consortium. RePORT links national tuberculosis cohorts—initially in India, Brazil, and South Africa, with expansion to additional sites—through a Common Protocol that standardizes clinical definitions, visit schedules, and the collection, quality assurance, and in-country stewardship of data and biospecimens. Each country builds and governs its own repositories while adhering to shared standards, so that specimens and datasets are interoperable and suitable for pooled analyses. This architecture enables capacity building in high-burden settings, supports South-South and South-North collaborations, and shortens the path from discovery to trial readiness for diagnostics, vaccines, and therapies. By design, RePORT operationalizes generosity: local ownership with shared rules for contribution, access, and credit, thereby creating a repeatable template for translational research networks beyond tuberculosis [7].

Importantly, this model also illustrates how generosity can address structural inequities in global science. By prioritizing in-country stewardship, capacity building, and equitable participation, RePORT challenges traditional models in which data and resources are extracted from high-burden settings and concentrated in institutions in the Global North. Instead, it supports a more distributed model of knowledge production, in which scientific leadership, authorship, and infrastructure development are shared across contexts. In this sense, generosity is not only a collaborative practice, but also a mechanism for redistributing visibility, credit, and opportunity within the global scientific ecosystem.

On a smaller scale, generosity often reveals itself in gestures that leave no trace in metrics or bibliometrics, such as sharing an idea over coffee, connecting a young colleague with a senior researcher, or giving space for others to present first. These acts might not count toward impact factors, but they shape scientific communities in profound ways. Science, at its best, is a collective endeavor, and generosity is the glue that keeps the collective from disintegrating into a pile of isolated egos.

## The Traditional Model of Academic Success

The dominant model of success in academia thrives on scarcity. It measures individuals through h-indexes, citations, and first- or last-authorships, all of which assume that recognition must be a zero-sum game [8]. In this system, your success is interpreted as someone else's failure, and the currency of prestige is accumulated by exclusion. The system rewards those who guard their data until the moment of publication, who compete for grants as if they were personal trophies, and who see colleagues less as collaborators than as potential rivals.

This model creates an atmosphere of anxiety and mistrust. Young scientists learn quickly that generosity is a liability. Share your idea too soon and someone else might publish it before you. Offer your dataset and you risk losing the first-author paper. Support a colleague's success too enthusiastically and you might find yourself overshadowed. The traditional model frames generosity as weakness and competition as strength, as if science were a gladiatorial contest rather than a collective search for truth.

There are undeniable achievements produced within this model. Competition drives ambition, and ambition produces results. Yet the costs are increasingly evident. Burnout, toxic lab cultures, plagiarism scandals, and inequities in access to funding all stem from the myth that science thrives when individuals fight each other for crumbs of recognition [9]. The model survives because it is convenient to bureaucracies and profitable to publishers, but it suffocates the very creativity it claims to reward. Actually, this is a dangerous trap: it is impossible to do science alone. And seeing rivalry where there should be collaboration is the key to this individual eventually finding themselves alone—and perhaps here, social and collaborative "failure" carries more weight than the journal cover that was published last month.

## Generosity as Power

Against this backdrop, generosity may appear as an act of resistance, but it is also an act of strategy. Generosity is not weakness—it is power. It transforms influence from something fragile and temporary into something enduring and expansive. A paper might earn your citations, but a culture of generosity creates networks of trust that outlast any citation count.

History provides striking examples. When Jonas Salk developed the polio vaccine, he refused to patent it, famously asking,



XSL•FO
RenderX

"Could you patent the sun?" [10]. That act of generosity ensured rapid global access and saved millions of lives. Similarly, the Human Genome Project embraced an ethos of open data sharing through the Bermuda Principles, requiring that sequence data be released within 24 hours [11]. This act of generosity neutralized private efforts to monopolize the human genome and revolutionized biomedical research. In addition, this approach operationalized generosity through data governance, enabling global participation and limiting the consolidation of knowledge within proprietary systems.

More recently, the COVID-19 pandemic showed how generosity scales. The unprecedented speed of vaccine development was made possible by rapid, open sharing of viral sequences [12]. The alternative—secrecy, proprietary data, nationalistic hoarding—would have prolonged the catastrophe. Generosity saved lives, while selfishness would have cost them. These examples suggest that, under certain conditions, openness and shared access can function as accelerators of discovery, particularly in contexts requiring coordinated, transnational action.

The examples discussed throughout this viewpoint are illustrative and were selected to clarify how generosity can operate across different scales of scientific practice. They are intended to support conceptual development and practical reflection rather than to constitute a systematic or scoping review of the literature. Across these cases, generosity operates through identifiable mechanisms: open access to data, equitable distribution of credit, and collaborative infrastructure that lowers barriers to participation. These mechanisms transform generosity from an individual disposition into a reproducible feature of scientific systems.

In my own career, I have seen how generosity amplifies influence across continents. When you mentor generously, your students carry your ethos into their own careers, spreading it in places you will never visit. When you share credit generously, collaborators are eager to return with new ideas and resources. When you share knowledge generously, your field grows stronger and your own work finds deeper resonance. Generosity multiplies success, while selfishness isolates it. Furthermore, when mentees are not recognized for their contributions, they may grow disenchanted with science not because of its true potential, but because of failures in mentorship. This neglect can dull the spark of curiosity and drive, pushing away brilliant minds who might have flourished if nurtured. In the end, selfishness not only isolates success; it risks silencing futures that could have expanded the field itself. The "takers" of science may enjoy brief moments of visibility, but their legacies are fragile, sustained only as long as they can monopolize attention.

## The Future of Science Through Generosity

Future generations face problems so vast that the old model of academic rivalry will collapse under their weight. Climate change, global pandemics, and the inequities of health cannot be solved by isolated laboratories fighting for primacy. These challenges demand collaborations that cross borders, disciplines, and generations [13]. For such collaborations to thrive, generosity must become the rule, not the exception.

Teaching young scientists to embrace generosity is not sentimentalism, it is survival. Rather than framing success solely in terms of individual achievement, mentorship and education can emphasize collaborative capacity, responsible data sharing, and equitable participation in knowledge production. In this context, generosity is not a form of self-sacrifice, but an investment in collective resilience—and generosity should extend beyond academic walls, whether in health care services, efforts to reduce economic and access inequalities, or general well-being [14-16].

When scientific activity is pursued primarily for individual recognition, concerns about priority and visibility may constrain collaboration and slow the circulation of knowledge. In contrast, when the purpose is to nourish science, empowering others is not a threat but a natural extension of one's mission. By expanding capacity and enabling new voices, the mentor amplifies the very contribution they set out to make, ensuring it reverberates far beyond individual achievements.

Institutions, too, must change. If universities and funding agencies continue to reward individual accumulation above collective impact, they will train generations of scientists ill-prepared for the realities ahead. Instead, we need metrics that value team science [17], platforms that encourage data sharing, and recognition systems that highlight mentorship and collaboration. Generosity must be made visible, measurable, and desirable.

In addition, emerging forms of digital publishing and open science infrastructures create new opportunities to make generosity more visible and traceable. Expanded reference lists, open peer review, data citation practices, and alternative metrics allow for broader recognition of intellectual contributions that extend beyond traditional authorship. These developments suggest that, rather than remaining invisible, many forms of generous scientific practice can be formally acknowledged and integrated into evaluation systems, supporting a more inclusive and representative scholarly ecosystem.

## Framework for Incorporating Generosity Into Science

### Domains and Principles

To translate generosity from a utopian value into a reproducible characteristic of scientific practice, this article proposes a framework organized into three domains: research planning, governance, and evaluation. Within each domain, generosity should be operationalized through fundamental principles, associated practices, and metrics that can help guide implementation (readjusting whenever necessary). Across all domains and areas of knowledge, four interrelated principles underpin this framework: (1) open science (facilitating the timely and responsible sharing of data, methods, and knowledge), (2) equitable recognition (ensuring fair recognition of contributions across all roles, career stages, and geographic regions), (3) collaborative planning (structuring research



XSL•FO
RenderX

questions and processes to allow for meaningful participation across disciplines, expertise, and contexts), and (4) management (maintaining responsible, long-term care of data, resources, and scientific relationships).

Scientific participation and recognition are shaped by structural inequities related to geography, language, institutional resources, gender, and race. As a result, the implementation of generosity must explicitly account for these asymmetries. Without such consideration, practices intended to promote openness and collaboration may inadvertently reproduce existing disparities rather than mitigate them.

## Operational Domains

### *Research Design*

Generosity in research design involves incorporating external collaboration and accessibility into the study. This includes integrating data-sharing plans into study protocols, identifying potential multisite or interdisciplinary collaborating groups, and anticipating the later reuse of data, samples, and materials. Indicative metrics may include a minimum number of prespecified data-sharing plans, the number of collaborating institutions, and the reuse of datasets in secondary analyses.

### *Governance*

Governance structures determine how resources, decisions, and recognition are distributed within scientific collaborations. In this domain, generosity also requires actively paying attention to global and structural inequities, including ensuring that researchers from historically underrepresented regions and groups are not limited to roles of data collection but are meaningfully included in leadership, authorship, and decision-making processes.

Generous governance includes transparent authorization and contribution criteria; shared decision-making processes; and equitable access to data and materials. Support for capacity building (infrastructure, resources, analysis, computing power, among others), particularly in resource-limited contexts. Indicative metrics may include the inclusion of early-career researchers in leadership roles and documented policies for data access and recognition attribution.

### *Evaluation*

Evaluation systems shape which behaviors are rewarded in science. Incorporating generosity into evaluation requires recognition of collaborative outcomes (eg, shared datasets, consortium work) and clear inclusion of team contributions to the work, funding, and awards received by the group, among others. Indicative metrics may include citations of shared datasets; participation in multicenter studies; documented contributions beyond first/last authorship; number of students at different levels of knowledge present in each article; and metrics of representation of gender, race/ethnicity, and areas of academic training.

To operationalize generosity, institutions and research groups can adopt practical measures such as (1) incorporating data-sharing plans and timelines into study designs; (2) establishing transparent authorship and credit allocation criteria; (3) recognizing mentorship and collaborative contributions in promotion and funding decisions; and (4) developing indicators that capture collaborative outputs, such as shared datasets, multisite studies, and cross-disciplinary publications.

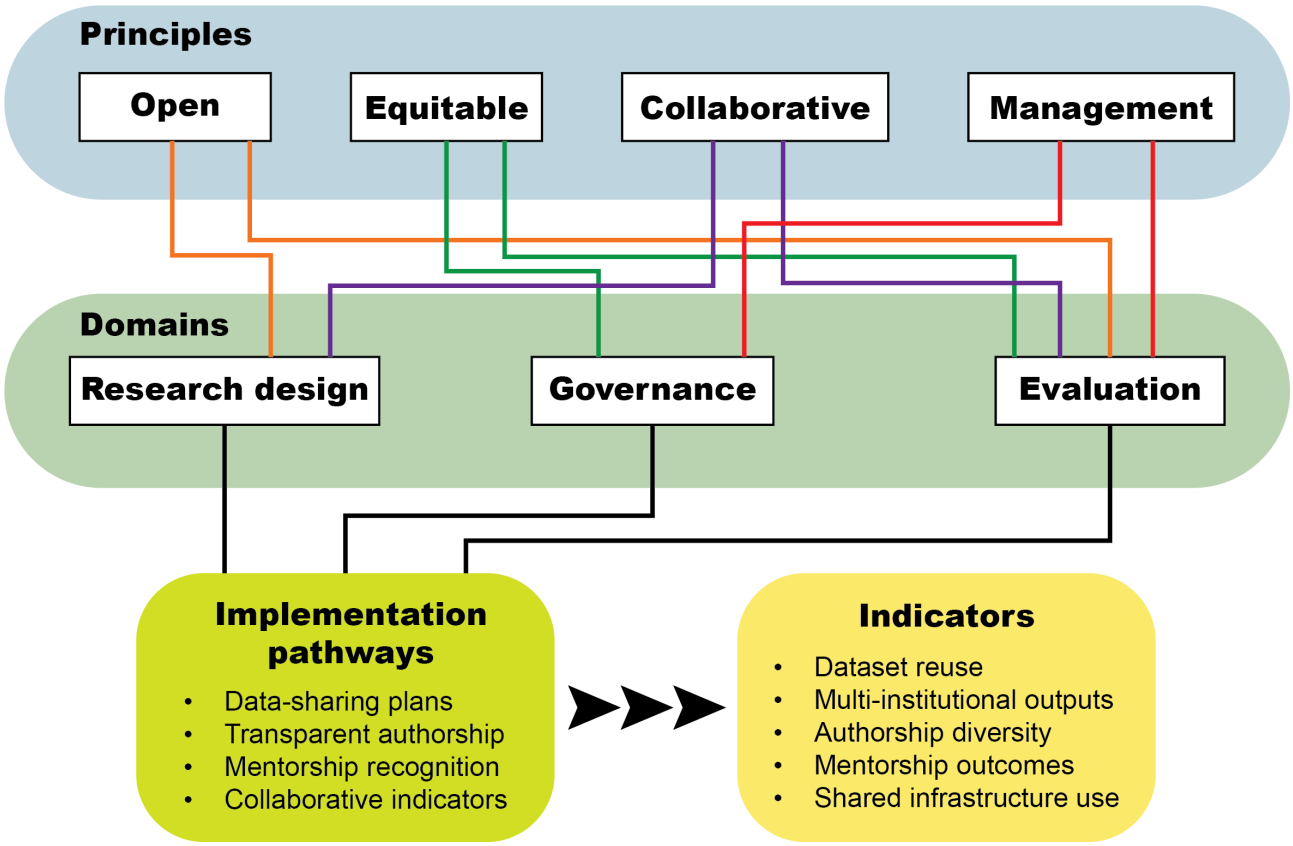
## Integration

Together, these elements position generosity not as an abstract ideal, but as an actionable and evaluable dimension of scientific practice (Figure 1). By integrating principles, domains, and indicators, this framework provides a structured approach for aligning collaborative values with the operational realities of research systems. Importantly, it allows generosity to be recognized, implemented, and assessed across different contexts, from individual projects to large-scale institutional initiatives. In doing so, it lays the groundwork for a shift from competition-centered models toward more collaborative, equitable, and sustainable forms of knowledge production.



XSL•FO
RenderX

**Figure 1.** Operational framework for embedding generosity in science. This figure illustrates a multilevel framework for embedding generosity in scientific practice. Core principles (top) inform three operational domains (middle), which are translated into actionable implementation pathways (bottom). Indicators provide mechanisms for evaluation and integration into research systems.




### Limitations

The framework proposed here is conceptual and intended to guide reflection and implementation rather than prescribe a universal model. Its adoption may face important constraints, including intellectual property considerations, regulatory requirements, time-sensitive research environments, and institutional evaluation systems that continue to privilege individual productivity over collective contribution. In some contexts, data sharing may need to be balanced against privacy, biosafety, sovereignty, or ethical concerns. Moreover, the meaning and feasibility of generosity may vary across disciplines, funding structures, and national settings. These boundary conditions do not diminish the value of the framework, but they do indicate that its implementation must be context-sensitive and institutionally supported.

## *Conclusion: A Call to Infectious Generosity*

This viewpoint has argued that generosity can be understood not only as a scientific virtue, but as a methodological orientation that shapes how research is designed, governed, and evaluated. It guides how we design collaborations, how we mentor, how we share, and how we dream. By aligning practices such as data sharing, equitable credit allocation, and collaborative infrastructure with core scientific goals, it resists the scarcity model and replaces it with abundance, not in the sense of infinite resources, but in the sense of infinite possibilities created through trust.

The framework proposed here highlights how generosity operates across multiple levels (from individual mentorship practices to institutional incentive structures) and identifies actionable pathways for its implementation. At the same time, its adoption is not without challenges, particularly in contexts shaped by competition for funding, intellectual property constraints, and entrenched evaluation systems.

Choosing generosity is not easy in a system that still worships rivalry. It requires courage to give credit when it could have been yours, to share data when secrecy seems safer, to open doors when others are closing them. Yet courage is exactly what science demands of us, and generosity is the most courageous act of all, because it asks us to trust that others will carry forward what we began.

The future of science will be written by those who dare to be generous. If generosity becomes infectious, if it spreads through laboratories and institutions, we may finally build a culture that enables science to deliver on its promises. Not a race, not a war, but a collective journey toward understanding and healing. What kind of research ecosystem do we want to leave for the next generation? One built on rivalry, or one built on generosity? The answer will determine not just the success of our careers, but the fate of our species.



XSL•FO
RenderX

## Acknowledgments

The author thanks Dr Mariana Araújo-Pereira, PhD, from the Oswaldo Cruz Foundation, Brazil, for her generous contributions, collaboration, and critical review of the manuscript. The author declares the use of generative AI (GAI) according to the GAIDeT taxonomy (2025). The following task was delegated to a GAI tool under full human supervision: summarizing text. The GAI tool used was ChatGPT-4.5. Responsibility for the final manuscript lies entirely with the author.

## Funding

The author declared no financial support was received for this work.

## Conflicts of Interest

None declared.

## References

1. Anderson C. Infectious Generosity: The Ultimate Idea Worth Spreading: Random House; 2024.
2. Davies SW, Putnam HM, Ainsworth T, et al. Promoting inclusive metrics of success and impact to dismantle a discriminatory reward system in science. PLoS Biol 2021 Jun;19(6):e3001282. [doi: 10.1371/journal.pbio.3001282] [Medline: 34129646]
3. Ray KS, Zurn P, Dworkin JD, Bassett DS, Resnik DB. Citation bias, diversity, and ethics. Account Res 2024 Dec;31(2):158-172. [doi: 10.1080/08989621.2022.2111257] [Medline: 35938378]
4. Macgilchrist F, Potter J, Williamson B. Reading internationally: if citing is a political practice, who are we reading and who are we citing? Learn Media Technol 2022 Oct 2;47(4):407-412. [doi: 10.1080/17439884.2022.2140673]
5. Rucavado Rojas D, Postigo JC. The cycle of underrepresentation: structural and institutional factors limiting the representation of Global South authors and knowledge in the IPCC. Clim Change 2025 Feb;178(2):19. [doi: 10.1007/s10584-025-03857-z]
6. Holman L, Stuart-Fox D, Hauser CE. The gender gap in science: how long until women are equally represented? PLoS Biol 2018 Apr;16(4):e2004956. [doi: 10.1371/journal.pbio.2004956] [Medline: 29672508]
7. Hamilton CD, Swaminathan S, Christopher DJ, et al. RePORT International: advancing tuberculosis biomarker research through global collaboration. Clin Infect Dis 2015 Oct 15;61Suppl 3(Suppl 3):S155-S159. [doi: 10.1093/cid/civ611] [Medline: 26409277]
8. Ioannidis JPA. Why most published research findings are false. PLOS Med 2005 Aug;2(8):e124. [doi: 10.1371/journal.pmed.0020124] [Medline: 16060722]
9. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A 2005 Nov 15;102(46):16569-16572. [doi: 10.1073/pnas.0507655102] [Medline: 16275915]
10. Oshinsky DM. Polio: An American Story: Oxford University Press; 2006.
11. Marshall E. Bermuda rules: community spirit, with teeth. Science 2001 Feb 16;291(5507):1192. [doi: 10.1126/science.291.5507.1192] [Medline: 11233433]
12. Besançon L, Peiffer-Smadja N, Segalas C, et al. Open science saves lives: lessons from the COVID-19 pandemic. BMC Med Res Methodol 2021 Jun 5;21(1):117. [doi: 10.1186/s12874-021-01304-y] [Medline: 34090351]
13. Bryant JH, Harrison PF, Institute of Medicine (US) Board on International Health. Global Health in Transition: A Synthesis: Perspectives from International Organizations: National Academies Press; 1996. [Medline: 24967490]
14. Montori VM. Turning away from industrial health care toward careful and kind care. Acad Med 2019 Jun;94(6):768-770. [doi: 10.1097/ACM.0000000000002534] [Medline: 30475268]
15. Tsiotsou RH, Seger-Guttmann T. The generosity ecosystem framework (GEF): a comprehensive approach to giving. Int Rev Public Nonprofit Mark 2025 Sep;22(3):489-515. [doi: 10.1007/s12208-025-00435-y]
16. Wang MC, Tran KK, Nyutu PN, Fleming E. Doing the right thing: a mixed-methods study focused on generosity and positive well-being. Journal of Creativity in Mental Health 2014 Jul 3;9(3):318-331. [doi: 10.1080/15401383.2014.890556]
17. Hall KL, Vogel AL, Huang GC, et al. The science of team science: a review of the empirical evidence and research gaps on collaboration in science. Am Psychol 2018;73(4):532-548. [doi: 10.1037/amp0000319] [Medline: 29792466]

## Abbreviations

**RePORT:** Regional Prospective Observational Research in Tuberculosis



XSL•FO
RenderX

*Edited by M Balcarras; submitted 27.Sep.2025; peer-reviewed by C Craven, S Mohanadas; revised version received 29.Apr.2026; accepted 30.Apr.2026; published 18.Jun.2026.*

*Please cite as:*
*Andrade BB*
*Generosity as a Scientific Method: Building Knowledge and Community in a Competitive World*
*Interact J Med Res 2026;15:e84933*
*URL: https://www.i-jmr.org/2026/1/e84933*
*doi:10.2196/84933*

© Bruno B Andrade. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 18.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Rethinking Basic Science Teaching in Medical Education: A Viewpoint on Its Past, Present, and Future

Kátia M Avena[1,2,3], PhD; Ana Paula A Brito[4,5], PhD; Luiz F Quintanilha[5,6], PhD; Mariana Araújo-Pereira[1,2], PhD; Bruno B Andrade[1,2,3,7,8], PhD

[1]Instituto Monster de Ensino, Assistência, Pesquisa e Desenvolvimento Tecnológico em Saúde, Av. Luís Viana Filho, 3230Imbuí, Salvador, Bahia, Brazil

[2]Clinical and Translational Research Laboratory / Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Bahia, Brazil

[3]Academic Operations, Clariens Educação, Salvador, Bahia, Brazil

[4]Universidade do Estado da Bahia, Salvador, Bahia, Brazil

[5]School of Medicine, Faculdade Zarns Salvador, Salvador, Bahia, Brazil

[6]Graduate Program in Regional and Urban Development, Salvador University, Salvador, Brazil

[7]Division of Infectious Diseases, Johns Hopkins Medicine, Baltimore, MD, United States

[8]Department of International Health, Bloomberg School of Public Health, Johns Hopkins Medicine, Baltimore, MD, United States

**Corresponding Author:**
Kátia M Avena, PhD
Instituto Monster de Ensino, Assistência, Pesquisa e Desenvolvimento Tecnológico em Saúde, Av. Luís Viana Filho, 3230Imbuí, Salvador, Bahia, Brazil

## Abstract

Basic sciences remain a cornerstone of medical education, but the terms of the debate about their importance have changed. The central risk is no longer their formal exclusion from medical curricula but their curricular invisibility: the gradual dilution of mechanistic reasoning within integrated, technology-mediated, and institutionally unequal training environments. In this Viewpoint, we revisit the past, present, and future of basic science teaching to argue that reform must protect conceptual progression, curricular visibility, and clinical transfer. Active learning, curricular integration, digital tools, and AI may strengthen scientific literacy, but only when they are anchored in faculty development, institutional support, and explicit commitments to rigor and equity.

***(Interact J Med Res 2026;15:e80646)*** doi:10.2196/80646

**KEYWORDS**

medical education; medical schools; medical training; basic sciences; technology; pedagogy

## Introduction

The teaching of basic sciences has long occupied a central place in medical education, providing the theoretical basis for clinical reasoning and professional practice [1]. Disciplines such as anatomy, biochemistry, and physiology have traditionally supported medical training by helping students understand the structure and function of the human body in health and disease [2,3]. In this article, the term “basic sciences” refers to biomedical disciplines that provide the conceptual and mechanistic basis for understanding normal and pathological processes and for interpreting clinical phenomena. Given their historical role as a core component of medical curricula, revisiting how these disciplines are taught has become increasingly necessary, particularly in light of contemporary educational and health care challenges [2-4]. This discussion is especially important in the current context of curricular reform, in which efforts to improve relevance, integration, and efficiency may also, sometimes unintentionally, weaken the visibility, coherence, or depth of foundational scientific knowledge.

This challenge is particularly evident across many settings in the Global South, where the rapid expansion of medical education has often occurred alongside resource constraints, uneven institutional capacity, and, in some contexts, increasing privatization, raising concerns about training quality and professional preparedness [5-7]. In such settings, the central issue is no longer whether basic sciences remain important but how to preserve their role within increasingly complex, technology-mediated, and unequal educational settings. At the same time, the continuous evolution of medical knowledge and the growing complexity of health care have required substantial changes in how these disciplines are taught [3]. The task is not simply to modernize teaching but to do so without weakening the scientific foundations of medical training. As curricula move toward integration, competency-based education, and digital innovation, basic sciences may either be meaningfully recontextualized or progressively marginalized. This marginalization does not necessarily mean that basic sciences are formally removed from the curriculum. More often, it occurs when they remain present but become fragmented, are



XSL•FO
RenderX

superficially addressed, and are less visible as the conceptual basis for clinical reasoning within integrated curricula. This tension lies at the center of contemporary medical education: how to make the teaching of basic sciences more relevant, practical, and student-centered while preserving the rigor, depth, and conceptual coherence that sustain clinical reasoning and professional judgment.

This Viewpoint is a critical reflection grounded in institutional experience and supported by selected literature. Its purpose is not to review every model of basic science teaching but to clarify the conditions under which reform becomes educationally meaningful. Integration, active learning, and digital innovation matter only when they make the scientific basis of clinical reasoning more visible, more usable, and more durable throughout the curriculum.

The contribution of this article is therefore conceptual. We argue that the main threat to basic science education is not its absence from medical curricula but rather its curricular invisibility. Contemporary debates do not usually call for the formal removal of basic sciences from medical curricula. More often, they question the traditional basic-clinical distinction [3] or advocate the stronger dissolution of the preclinical phase into clinically oriented teaching [8]. The concern lies neither in integration itself nor in efforts to align basic sciences with clinical practice. On the contrary, such initiatives can strengthen medical education when they are well designed. The problem arises when reform weakens the conceptual progression of basic sciences, their curricular visibility, and their interpretive role within integrated curricula. This risk becomes more pronounced when technology is adopted more rapidly than pedagogy, when faculty development is treated as optional, and when educational expansion occurs without commensurate investment in infrastructure and quality assurance. The past-present-future structure is used here to show how this problem emerged, how it currently appears, and what must be preserved if reform is to strengthen rather than weaken the foundational and scientific training of physicians.

## The Past: The Historical Legacy of Basic Science Teaching in Medical Education

The modern place of basic sciences was consolidated during the broader transformation that made medical education more scientific, structured, and institutionally grounded [9-11]. The Flexner report [11] and the Dawson report [12] matter here less as historical landmarks than as sources of a tension that is still present: medical education gained scientific legitimacy, but the curricular form used to protect that legitimacy often separated mechanism from clinical use.

The Flexner report [11] strengthened the expectation that clinical practice should rest on foundational scientific knowledge [2,13,14]. Yet the curriculum that followed was commonly lecture-based, discipline-centered, and dependent on memorization and laboratory instruction. The result was productive and limiting at the same time. It gave basic sciences authority, but it also helped place them in a preclinical space that students could experience as distant from patient care. The Dawson report [12], by emphasizing the relation between education, service, and clinical context, pointed to a different possibility, but the compartmentalized model remained dominant.

Throughout much of the 20th century, basic science teaching remained organized around passive transmission, delayed application, and limited clinical contextualization [15,16]. This approach helped consolidate foundational knowledge, but it also encouraged a form of learning in which detail could accumulate without becoming reasoning. The problem became visible across anatomy, biochemistry, and microbiology, where students often encountered scientific content before they understood why it mattered for diagnosis, prognosis, or therapeutic judgment [17-22].

The historical lesson is therefore precise. The basic sciences were protected by a model that also isolated them. Current reforms still carry this inheritance: they must reduce fragmentation and excessive content load without dissolving the conceptual progression that allows foundational knowledge to become clinical reasoning [23]. The task is not to move away from science but to move scientific explanation closer to the work of interpretation, judgment, and care.

## The Present: Current Practices and Challenges in Basic Science Education

Basic science teaching in medical education is currently marked by both important advances and persistent tensions. In many schools, disciplines such as anatomy, physiology, biochemistry, pharmacology, and microbiology are being restructured to strengthen the relationship between foundational knowledge and clinical application. This movement reflects broader efforts to move away from fragmented, content-heavy models and toward curricula that are more integrated, interactive, and responsive to contemporary health care needs. However, this transformation remains incomplete, especially in settings where infrastructure is limited, institutional support is weak, or resistance to pedagogical change remains strong.

A central feature of this transition has been the attempt to overcome the longstanding divide between theory and practice. Traditional models, heavily centered on content accumulation and memorization, often produced cognitive overload and limited students' ability to connect foundational concepts with clinical reasoning [8,24]. In response, many medical schools have adopted more integrated approaches, including problem-based learning (PBL), case-based teaching, and spiral curricula, all of which seek to connect basic and clinical sciences more meaningfully [13,25,26]. Recent work in physiology education has also reinforced the value of active learning strategies in basic science teaching [27]. At the same time, curricular reforms have increasingly emphasized the horizontal and vertical integration of the basic sciences across the curriculum, supported by active learning strategies and faculty development efforts aimed at strengthening integration between basic and clinical sciences [28].

Clinically oriented integration may also support retention. Earlier clinical exposure, clerkship-based reinforcement, and



XSL•FO
RenderX

integrated assessment have been associated with stronger preservation of foundational knowledge [28,29]. In a recent single-institution study from Sudan, students performed better on integrated questions than on discipline-specific items, suggesting that basic science knowledge is more durable when students must use it to interpret clinical problems [29]. This evidence supports competency-based education, provided that competence is understood as reasoning with science rather than merely performing clinical tasks [30,31].

Even so, progress remains uneven. Many institutions continue to struggle with insufficient pedagogical support and limited investment in faculty development [32,33]. Educators trained in more traditional systems may find it difficult to adopt active methods, integrate theory with clinical reasoning, or use digital tools effectively [34-36]. Evidence also suggests that even when faculty value active learning more highly, traditional lectures often remain part of their routine, highlighting a possible mismatch between pedagogical beliefs and teaching practices [36]. This gap between curricular intention and classroom reality shows that reform depends not only on curricular redesign but also on preparing teachers to interpret and enact new models consistently. These barriers are especially relevant in contexts where nonphysician educators are responsible for much of the teaching in the basic sciences. Although these faculty members often bring strong disciplinary knowledge and considerable pedagogical potential, they may receive limited institutional recognition or support, which can hinder innovation and interdisciplinarity [37].

These tensions reveal a broader challenge in contemporary medical education: how to modernize teaching without weakening scientific rigor. Making basic science education more relevant does not mean reducing its depth but rather teaching it in ways that are more contextualized, clinically meaningful, and educationally effective. Real clinical cases, cross-disciplinary dialogue, and stronger collaboration between basic and clinical faculty can help students understand how foundational knowledge informs medical judgment and patient care. In this sense, the goal is not to replace the basic sciences but to reposition them more effectively within the curriculum.

Interdisciplinarity has become one of the clearest expressions of this repositioning. Contemporary curricula increasingly incorporate social sciences, public health, and ethics into the teaching of basic disciplines, reflecting the complexity and interconnectedness of modern health care [8,38]. This broader framework supports more patient-centered and socially responsive training while also helping students connect scientific knowledge with the human, ethical, and systemic dimensions of practice. Emerging approaches such as STEAM (science, technology, engineering, arts, and mathematics) may further enrich this process by encouraging interdisciplinary projects, creativity, empathy, and visual reasoning [39]. The approach's emphasis on collaboration, design, and real-world problem-solving can help humanize scientific training while preserving rigor. In medical education, this approach is particularly valuable because it reinforces the idea that science and human experience should not be taught as separate domains.

Technological innovation has also reshaped current teaching practices [40-42]. After the COVID-19 pandemic, digital tools became even more prominent in health professions education [43,44]. In anatomy, cadaveric dissection is now frequently complemented by 3D models, virtual dissections, imaging-based simulations, and, in some contexts, virtual reality [17,18]. These tools can expand access, reduce dependence on costly physical resources, and offer flexible opportunities for visualization and repetition. Although cadaveric dissection retains important symbolic and educational value, newer strategies have broadened the possibilities for anatomy teaching. Of note, technology is most effective when it supports conceptual understanding, guided interpretation, and clinical transfer rather than functioning merely as a substitute for traditional materials or as an isolated innovation.

Similar changes are evident in other disciplines. In histology, pathology, and microbiology, virtual laboratories and digital microscopy have shown educational value while also reducing ethical and logistical concerns related to animals and biological materials [45-49]. In biochemistry, a field often perceived as abstract, gamification, simulation, and clinically oriented cases have made learning more engaging and relevant [50,51]. These innovations make learning more meaningful. Given the direct clinical implications of biochemical knowledge, active and contextualized approaches are particularly important in this area [52,53]. Physiology has also evolved through the use of digital simulations and gamified environments that connect normal and abnormal function while reducing reliance on animal-based teaching in nonclinical settings [54,55]. Across these examples, the underlying issue is not simply the adoption of new tools but the effort to make foundational scientific knowledge more accessible, meaningful, and applicable.

Yet these innovations do not affect all institutions equally. The expansion of online and technology-enhanced education has increased flexibility, autonomy, and opportunities for personalized learning, but it has also deepened inequalities between institutions and regions [7,56]. Many schools, especially in low- and middle-income countries (LMICs), still lack reliable internet access, adequate equipment, or sufficient technical and pedagogical support. This digital divide threatens equity and inclusion and reinforces the idea that innovation without investment may widen rather than reduce disparities [4].

Faculty development is therefore not an accessory to reform; it is the condition that makes reform possible [57]. Teachers need support to design active learning, construct assessments that reward reasoning, connect basic mechanisms to clinical decisions, and use digital tools without allowing the tool to become the lesson [37]. AI literacy now belongs in this preparation, but it should be framed critically: educators must learn where AI can reduce routine work and where human judgment, feedback, and mentorship remain irreplaceable. This support must include physician and nonphysician educators because both carry the responsibility for making science visible within medical training.

Structural problems also remain highly relevant. In many settings in the Global South, medical schools continue to face



XSL•FO
RenderX

curricular fragmentation, inadequate infrastructure, faculty shortages, and uneven regulatory capacity, all of which compromise training quality [5,6,58]. In some contexts, these challenges have been further intensified by the rapid expansion of medical education without proportional investment in educational resources, faculty support, or quality assurance mechanisms. Increasing privatization, weak regulation, and poor coordination between education and health care systems have further widened disparities and left many curricula insufficiently aligned with real health care needs [6,7,58-61].

The present moment is therefore defined by a difficult coexistence. Innovation is real, but so is inertia. Integration, interdisciplinarity, and technology have improved parts of basic science education, while structural weakness still determines how far these gains can extend. Preserving the basic sciences now requires more than new tools or new language. It requires institutional choices that protect scientific depth, make foundational knowledge clinically actionable, and reduce the gap between well-resourced and underresourced schools.

## The Future: Emerging Trends and Future Perspectives in Basic Science Education

The future of basic science education will be shaped less by the availability of new technologies than by the judgment used to place them within the curriculum. AI, augmented reality, virtual reality, and high-fidelity simulation can make learning more adaptive and immersive. They can also create another layer of distraction if they are adopted as symbols of modernization rather than as instruments of understanding. Future innovation must therefore be measured by a simple standard: whether it helps students reason more clearly from mechanism to clinical consequence. Concerns about digital fatigue and screen-mediated disconnection make this standard even more important, since medical education still depends on attention, dialogue, mentorship, and relational learning [62,63].

AI offers important possibilities for personalized education, adaptive learning, and scalable instructional support in medical training [64]. Adaptive systems can analyze student performance in real time, provide targeted feedback, and help address knowledge gaps more quickly. These functions may improve learning efficiency. However, the role of AI in replacing teaching competencies remains uncertain. A recent study with educators and technology experts found agreement only in specific areas, particularly those involving low-complexity tasks such as preparing materials or grading [65]. In contrast, linking theory to practice, anticipating student needs, and adjusting strategies were seen as activities that still depend on human judgment. Empathy and critical insight were also identified as irreplaceable aspects of teaching.

These findings suggest that AI can support repetitive tasks but not replace human guidance. Medical education continues to require emotional awareness, empathy, and active listening [66]. At a time when concerns about screen fatigue and digitally mediated disconnection are becoming more visible, preserving human interaction in educational settings becomes even more important [62,63]. In this context, the educator's role is shifting from content delivery to learning facilitation. Teachers are increasingly expected to act as mentors who guide critical thinking and ethical reflection while helping students use technology wisely and maintain human values at the center of education [66]. The educator of the future will therefore need to combine disciplinary expertise, pedagogical judgment, digital literacy, and the ability to sustain reflective and relational forms of learning in increasingly technology-rich environments.

Virtual and augmented reality are also reshaping learning. These technologies create highly interactive environments in which students can explore anatomy and physiology through 3D simulations. Such experiences can improve spatial reasoning and decision-making in clinical scenarios while allowing learners to practice safely before engaging with real patients. However, their expansion should complement, rather than displace, the relational dimensions of medical education, particularly in areas where communication and professional identity depend on interaction with teachers, peers, and patients [67].

Future curricula will likely continue to emphasize collaboration, interdisciplinarity, and competency-based learning. This broader scope can bring ethics, social sciences, and digital health closer to basic sciences, but only if integration preserves the logic of scientific explanation. Memorization should give way to application, reflection, and adaptability, yet application without a foundation becomes fragile. The value of basic sciences lies in their capacity to support reasoning, interpretation, and decision-making across clinical and social contexts rather than in the isolated accumulation of facts.

These changes point to a broader structural transformation. Basic science education is moving from rigid and fragmented models toward curricula that integrate disciplines, technology, and human development. Table 1 summarizes these shifts across pedagogical approaches, curricular structure, teaching tools, engagement, assessment, and clinical integration. The table should be read as a map of pressures rather than a prediction of automatic progress. Equity will decide whether these innovations become shared educational gains or another mechanism of stratification. In low- and middle-income settings, aspirations toward innovation often coexist with unstable connectivity, limited simulation resources, constrained faculty capacity, and uneven regulatory support.



XSL•FO
RenderX

**Table .** Summary comparison of past, present, and future trends in basic science teaching in medical education.

| Aspects | Past trends | Present trends | Future trends |
|---|---|---|---|
| Pedagogical approach | Teacher-centered lectures and passive learning | Student-centered active learning, including problem-based learning and team-based learning | Personalized and adaptive learning used to strengthen reasoning and feedback |
| Curricular structure | Isolated disciplines with minimal integration across areas | Integrated, system-based, and interdisciplinary curricula | Competency-based curricula with explicit progression of basic science concepts |
| Teaching tools | Textbooks, cadaveric dissection, and chalkboards | Cadaveric dissections with virtual reality, augmented reality, online platforms, and simulations | Immersive simulation; AI-supported tutoring; and science, technology, engineering, arts, and mathematics–informed resources tied to clear learning goals |
| Student engagement | Passive learning focused on memorization | Active participation, problem-solving, and collaborative learning | Autonomous learning paths supported by guided feedback and clinically grounded problem-solving |
| Assessment methods | High-stakes examinations and rote memorization | Formative assessments, objective structured clinical examinations, and competency-based evaluations | Continuous assessment of reasoning, knowledge transfer, and reflective clinical application |
| Integration with clinical practice | Limited and delayed clinical application | Early clinical exposure and integration of clinical cases | Longitudinal integration of basic and clinical sciences while preserving conceptual visibility |

Faculty development will remain critical in this future landscape [68]. Advanced technologies will not improve learning by themselves. Educators must know how to translate complex disciplinary content into clinical meaning, how to design feedback around reasoning, and how to decide when a digital tool adds value and when it distracts from the concept being taught. In basic science education, sustainable reform depends on this pedagogical judgment.

Students will also need to adapt. Active learning and technology-based education require new attitudes toward learning. Students must take greater responsibility for their own progress and develop autonomy, critical thinking, and resilience [69,70]. Success in this context depends on their ability to seek knowledge independently and apply it in dynamic situations. Consequently, the focus shifts from passive absorption to active exploration, supported by strategies that encourage self-regulation and lifelong learning. On the other hand, expectations of autonomy should be balanced with adequate guidance, feedback, and structured support, since students differ considerably in their readiness for self-directed learning and in their ability to navigate complex educational environments.

At the same time, the future of basic science education is threatened by erosion that can be difficult to see. Course hours may shrink, practical experiences may disappear, research requirements may be removed, and student-faculty ratios may grow while the curriculum still appears formally integrated [71]. The consequence is a thinner scientific literacy. Innovation will produce lasting benefit only if it protects the time, faculty, and pedagogical structure required for students to understand mechanisms deeply enough to use them.

A sustainable future for basic science education depends on political will, equitable investment, and cultural change within institutions. Technology can help, but it cannot compensate for weak pedagogy, undervalued faculty, or inadequate infrastructure. The central challenge is to modernize teaching while protecting the scientific and human foundations of physician training across diverse educational realities.

## Conclusions

Basic science education is being transformed by integration, active learning, and technology. These changes can improve medical training, but they can also make foundational knowledge less visible if reform is guided by novelty rather than by educational purpose. Across the past, present, and future of medical education, the central question is no longer whether basic sciences matter. It is whether medical schools can preserve the conceptual progression through which anatomy, physiology, biochemistry, microbiology, pharmacology, and pathology become clinical reasoning.

Defending the basic sciences does not mean returning to isolated preclinical courses or resisting curricular integration. It means refusing a weaker bargain in which science remains present in name but loses depth, sequence, and interpretive power. Integration is valuable when it makes mechanisms clinically intelligible. Technology is valuable when it deepens understanding rather than merely replaces attention. Faculty development is valuable when it equips teachers to connect explanation with judgment. The future of medical education will depend on this alignment: innovation that is strong enough to modernize training and scientific literacy that is strong enough to sustain the physicians who will practice in conditions of uncertainty.



XSL•FO
RenderX

## Acknowledgments

During manuscript preparation, the authors used ChatGPT (version GPT-5.6; OpenAI) to assist with language editing and formatting. The use of generative AI was limited to stylistic and linguistic support and did not involve data generation, analysis, interpretation, or independent modification of the scientific content. All AI-assisted outputs were critically reviewed and revised by the authors, who take full responsibility for the final content of the manuscript.

## Funding

This work was supported by the Intramural Research Program of Fundação Oswaldo Cruz (Fiocruz), Brazil. The funders had no role in the design of the study; in the collection, analysis, or interpretation of the literature; in the writing of the manuscript; or in the decision to submit the manuscript for publication.

## Authors' Contributions

Conceptualization: KMA, BBA
Data curation: BBA
Funding acquisition: BBA
Methodology: KMA, BBA
Project administration: BBA
Supervision: BBA
Writing—original draft: KMA, APAB, LFQ, BBA
Writing—review and editing: KMA, APAB, LFQ, MA-P, BBA

## Conflicts of Interest

None declared.

## References

1. Saleh SA, Adly HM. Integrating basic sciences and clinical practice: a cross-sectional study of UQUMED's medical education approach. Galician Med J 2024;31(1):e-GMJ2024-A02. [doi: 10.21802/e-GMJ2024-A02]
2. Finnerty EP, Chauvin S, Bonaminio G, Andrews M, Carroll RG, Pangaro LN. Flexner revisited: the role and value of the basic sciences in medical education. Acad Med 2010 Feb;85(2):349-355. [doi: 10.1097/ACM.0b013e3181c88b09] [Medline: 20107367]
3. Bandiera G, Boucher A, Neville A, Kuper A, Hodges B. Integration and timing of basic and clinical sciences education. Med Teach 2013 May;35(5):381-387. [doi: 10.3109/0142159X.2013.769674] [Medline: 23444888]
4. Frenk J, Chen L, Bhutta ZA, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. Lancet 2010 Dec 4;376(9756):1923-1958. [doi: 10.1016/S0140-6736(10)61854-5] [Medline: 21112623]
5. Mullan F, Frehywot S, Omaswa F, et al. Medical schools in sub-Saharan Africa. Lancet 2011 Mar 26;377(9771):1113-1121. [doi: 10.1016/S0140-6736(10)61961-7] [Medline: 21074256]
6. Sabde Y, Diwan V, Mahadik VK, et al. Medical schools in India: pattern of establishment and impact on public health - a geographic information system (GIS) based exploratory study. BMC Public Health 2020 May 24;20(1):755. [doi: 10.1186/s12889-020-08797-0] [Medline: 32448195]
7. Avena KM, Quintanilha LF, Luzardo Filho RL, Andrade BB. Lessons learned from the expansion of medical schools in Brazil: a review of challenges and opportunities. Front Educ 2024;9:1494445. [doi: 10.3389/feduc.2024.1494445]
8. Oktar S. Basic medical sciences should be mainly taught by clinicians for a tight integration of basic and clinical sciences in medical education. Acta Medica Alanya 2022;6(3):320-325. [doi: 10.30565/medalanya.1159175]
9. Cruse JM. History of medicine: the metamorphosis of scientific medicine in the ever-present past. Am J Med Sci 1999 Sep;318(3):171-180. [doi: 10.1016/S0002-9629(15)40609-3] [Medline: 10487407]
10. Modanlou HD. Historical evidence for the origin of teaching hospital, medical school and the rise of academic medicine. J Perinatol 2011 Apr;31(4):236-239. [doi: 10.1038/jp.2010.162] [Medline: 21233794]
11. Flexner A. Medical education in the United States and Canada. : The Carnegie Foundation for the Advancement of Teaching; 1910 URL: https://archive.carnegiefoundation.org/publications/pdfs/elibrary/Carnegie_Flexner_Report.pdf [accessed 2026-07-20]
12. Interim report on the future provision of medical and allied services 1920 (Lord Dawson of Penn). : United Kingdom Ministry of Health, Consultative Council on Medical and Allied Services; 1920 URL: https://tinyurl.com/4mbcefcw [accessed 2026-07-21]
13. Cooke M, Irby DM, Sullivan W, Ludmerer KM. American medical education 100 years after the Flexner report. N Engl J Med 2006 Sep 28;355(13):1339-1344. [doi: 10.1056/NEJMra055445] [Medline: 17005951]



XSL•FO
RenderX

14. Babker A. Impact of the Flexner's report on health professional education. World J Adv Res Rev 2023;19(3):939-941. [doi: 10.30574/wjarr.2023.19.3.1888]
15. Dolan S, Mallott DB, Emery JA. Passive learning: a marker for the academically at risk. Med Teach 2002 Nov;24(6):648-649. [doi: 10.1080/01421590216082] [Medline: 12623462]
16. Torralba KD, Doo L. Active learning strategies to improve progression from knowledge to action. Rheum Dis Clin North Am 2020 Feb;46(1):1-19. [doi: 10.1016/j.rdc.2019.09.001] [Medline: 31757278]
17. Darras KE, de Bruin AB, Nicolaou S, et al. Is there a superior simulator for human anatomy education? How virtual dissection can overcome the anatomic and pedagogic limitations of cadaveric dissection. Med Teach 2018 Jul;40(7):752-753. [doi: 10.1080/0142159X.2018.1451629] [Medline: 29569960]
18. McMenamin PG, McLachlan J, Wilson A, et al. Do we really need cadavers anymore to learn anatomy in undergraduate medicine? Med Teach 2018 Oct;40(10):1020-1029. [doi: 10.1080/0142159X.2018.1485884] [Medline: 30265177]
19. Huggins D, Goacher P. Using technology to teach anatomy and physiology. J Oper Dep Pract 2013 Nov;1(2):80-84. [doi: 10.12968/jodp.2013.1.2.80]
20. Santos VA, Barreira MP, Saad KR. Technological resources for teaching and learning about human anatomy in the medical course: systematic review of literature. Anat Sci Educ 2022 Mar;15(2):403-419. [doi: 10.1002/ase.2142] [Medline: 34664384]
21. Vieira LQ, Nicoli JR, Prado VF, Santoro MM, Teixeira SM. Teaching biochemistry: a practical approach [Article in Portuguese]. Rev Ensino Bioquím 2001;1(1):20-26 [FREE Full text] [doi: 10.16923/reb.v1i1.6]
22. Yokaichiya DK, Galembeck E, Torres BB. What do students from different courses look for in extracurricular biochemistry subjects [Article in Portuguese]. Rev Ensino Bioquím 2004;2(1):37-44 [FREE Full text] [doi: 10.16923/reb.v2i1.17]
23. Cooke M, Irby DM, O'Brien BC. Educating Physicians: A Call for Reform of Medical School and Residency: Jossey-Bass; 2010.
24. Ma CW. How to advance medical education using journal articles? Insight from problem-based learning. GMS J Med Educ 2022;39(4):Doc48. [doi: 10.3205/zma001569] [Medline: 36310889]
25. Hurtubise L, Hall E, Sheridan L, Han H. The flipped classroom in medical education: engaging students to build competency. J Med Educ Curric Dev 2015;2:JMECD.S23895. [doi: 10.4137/JMECD.S23895] [Medline: 35187252]
26. Wartman SA, Combs CD. Reimagining medical education in the age of AI. AMA J Ethics 2019 Feb 1;21(2):E146-E152. [doi: 10.1001/amajethics.2019.146] [Medline: 30794124]
27. Gaddas M, Ben Daly A, Ben Dhiab M, Ben Saad H. Rethinking active learning in medical education: a comparative study of inquiry-based and team-based learning on student performance and satisfaction. Adv Physiol Educ 2026 Mar 1;50(1):22-31. [doi: 10.1152/advan.00199.2025] [Medline: 41196126]
28. Norris ME, Cachia MA, Johnson MI, Martin CM, Rogers KA. Are clerks proficient in the basic sciences? Assessment of third-year medical students' basic science knowledge prior to and at the completion of core clerkship rotations. Med Sci Educ 2021;31(2):709-722. [doi: 10.1007/s40670-021-01249-3] [Medline: 34457921]
29. Khalid FA, Haroon HM, Eltom H, Ali AA. Retention of basic sciences knowledge at clinical years: a single institute study from Sudan. BMC Med Educ 2026 Jan 14;26(1):235. [doi: 10.1186/s12909-026-08599-9] [Medline: 41530731]
30. Schwartzstein RM, Rosenfeld GC, Hilborn R, Oyewole SH, Mitchell K. Redesigning the MCAT exam: balancing multiple perspectives. Acad Med 2013 May;88(5):560-567. [doi: 10.1097/ACM.0b013e31828c4ae0] [Medline: 23524933]
31. Prober CG, Heath C. Lecture halls without lectures--a proposal for medical education. N Engl J Med 2012 May 3;366(18):1657-1659. [doi: 10.1056/NEJMp1202451] [Medline: 22551125]
32. Quintanilha LF, Pereira MA, Andrade BB, Portilho EM. Teacher training and engagement in Brazilian health higher education: fragilities and strengths [Article in Portuguese]. Rev Bras Estud Pedagog 2023;104:e5293. [doi: 10.24109/2176-6681.rbep.104.5293]
33. Quintanilha LF, Farias CS, Andrade BB. Training and pedagogical involvement among higher education professors in health: an analysis of medical courses [Article in Portuguese]. Rev Inter Educ Sup 2020;7:e021026. [doi: 10.20396/riesup.v7i0.8659850]
34. Moen J, Shuck C. Medical educator teaching habits: origins, decisions, and strategies—a phenomenological inquiry (METHODS-PI). Discov Educ 2024;3:253. [doi: 10.1007/s44217-024-00362-w]
35. Oliveira TD, Medeiros RD, Pires MP, Pinto MD. Pedagogical preparation for physicians and their performance in the medical course [Article in Portuguese]. Rev Bras Educ Med 2018;42(3):171-177. [doi: 10.1590/1981-52712015v42n3rb2017066.r2ing]
36. Carvalho H, Dane FC, Whicker SA. Conceptions of learning and teaching for faculty who teach basic science. Med Sci Educ 2021 Mar;31(2):745-751. [doi: 10.1007/s40670-021-01264-4] [Medline: 34457923]
37. Quintanilha LF, Avena KM, Fatel QH, Jesus RF, Araújo-Pereira M, Andrade BB. Beyond the white coat: valuing the contributions of non-physician faculty in medical education in Brazil. Teach Learn Med 2025 Oct 17:1-9. [doi: 10.1080/10401334.2025.2569438] [Medline: 41104614]
38. Dahle LO, Brynhildsen J, Behrbohm Fallsberg M, Rundquist I, Hammar M. Pros and cons of vertical integration between clinical medicine and basic science within a problem-based undergraduate medical curriculum: examples and experiences from Linköping, Sweden. Med Teach 2002 May;24(3):280-285. [doi: 10.1080/01421590220134097] [Medline: 12098414]




XSL•FO
RenderX

39. Hardoim EL, Hardoim TF, Nakamura CR, Hardoim AH. Educação científica inclusiva: experiências interdisciplinares possíveis para o ensino de biologia e ciências naturais empregando o método STEAM [Article in Portuguese]. Lat Am J Sci Educ 2019;6:12056 [FREE Full text]
40. Hassall C, Lewis DI. Institutional and technological barriers to the use of open educational resources (OERs) in physiology and medical education. Adv Physiol Educ 2017 Mar 1;41(1):77-81. [doi: 10.1152/advan.00171.2016] [Medline: 28143826]
41. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. Acad Med 2024 Jan 1;99(1):22-27. [doi: 10.1097/ACM.0000000000005439] [Medline: 37651677]
42. Dhar P, Rocks T, Samarasinghe RM, Stephenson G, Smith C. Augmented reality in medical education: students' experiences and learning outcomes. Med Educ Online 2021 Dec;26(1):1953953. [doi: 10.1080/10872981.2021.1953953] [Medline: 34259122]
43. Jiang Z, Wu H, Cheng H, Wang W, Xie A, Fitzgerald SR. Twelve tips for teaching medical students online under COVID-19. Med Educ Online 2021 Dec;26(1):1854066. [doi: 10.1080/10872981.2020.1854066] [Medline: 33280546]
44. Jeffries PR, Bushardt RL, DuBose-Morris R, et al. The role of technology in health professions education during the COVID-19 pandemic. Acad Med 2022 Mar 1;97(3S):S104-S109. [doi: 10.1097/ACM.0000000000004523] [Medline: 34789662]
45. Marchevsky AM, Relan A, Baillie S. Self-instructional "virtual pathology" laboratories using web-based technology enhance medical school teaching of pathology. Hum Pathol 2003 May;34(5):423-429. [doi: 10.1016/s0046-8177(03)00089-3] [Medline: 12792914]
46. Bean LJ, Fridovich-Keil J, Hegde M, Rudd MK, Garber KB. The virtual diagnostic laboratory: a new way of teaching undergraduate medical students about genetic testing. Genet Med 2011 Nov;13(11):973-977. [doi: 10.1097/GIM.0b013e318225ac13] [Medline: 21811165]
47. Dee FR, Meyerholz DK. Teaching medical pathology in the twenty-first century: virtual microscopy applications. J Vet Med Educ 2007;34(4):431-436. [doi: 10.3138/jvme.34.4.431] [Medline: 18287469]
48. Blake CA, Lavoie HA, Millette CF. Teaching medical histology at the University of South Carolina School of Medicine: transition to virtual slides and virtual microscopes. Anat Rec B New Anat 2003 Dec;275(1):196-206. [doi: 10.1002/ar.b.10037] [Medline: 14628320]
49. Then SM, Kokolski M, Mbaki Y, Merrick D, Anderson S. An international collaborative approach to learning histology using a virtual microscope. Anat Histol Embryol 2023 Jan;52(1):21-30. [doi: 10.1111/ahe.12888] [Medline: 36373558]
50. Sait MS, Siddiqui Z, Ashraf Y. Advances in medical education and practice: student perceptions of the flipped classroom. Adv Med Educ Pract 2017;8:317-320. [doi: 10.2147/AMEP.S133328] [Medline: 28496377]
51. Surapaneni KM. "CARBGAME" (Card & Board Games in Medical Education) as an innovative gamification tool for learning clinical enzymology in biochemistry for first year medical students. Biochem Mol Biol Educ 2024;52(6):666-675. [doi: 10.1002/bmb.21857] [Medline: 39136227]
52. Emre HO, Karpuzoglu FH, Coskun C, et al. Utilization of biological variation data in the interpretation of laboratory test results - survey about clinicians' opinion and knowledge. Biochem Med (Zagreb) 2021 Feb 15;31(1):010705. [doi: 10.11613/BM.2021.010705] [Medline: 33380892]
53. Alfaqih MA, Khader YS, Bashir N, Nusair Z, Nuseir Q, Nusier M. Attitude of Jordanian physicians toward biochemistry and genetics. Biomed Res Int 2019;2019:3176951. [doi: 10.1155/2019/3176951] [Medline: 31815130]
54. Brazilian Guide for the Production, Maintenance, or Use of Animals for Teaching or Scientific Research Activities [Book in Portuguese]: Ministry of Science, Technology and Innovation, Brazil; 2023.
55. Maiche SD, Carneiro GS, Jesus TR, Brito AP, Avena KM, Quintanilha LF. Revolutionizing learning in medical education: an experience report on the use of gamification as a teaching strategy in practical classes of renal physiology [Article in Portuguese]. Diál Ciênc 2023;2(3):62-69 [FREE Full text] [doi: 10.7447/1678-0493.2022v2n3p62-69]
56. Fatel QH, Silva FG, Avena KM, Menezes RC, Andrade BB, Quintanilha LF. Shaping tomorrow's doctors: the impact of socioeconomic and institutional factors on medical education quality in Brazil: an ecological study. J Med Educ Curric Dev 2025;12:23821205251360521. [doi: 10.1177/23821205251360521] [Medline: 40757008]
57. McCoy L, Diaz SR, Dennis Baker S. The development of a faculty development program framework for medical education. J Med Educ Curric Dev 2026;13:23821205261439617. [doi: 10.1177/23821205261439617] [Medline: 42039676]
58. Tsinuel G, Tsedeke A, Matthias S, et al. Establishing medical schools in limited resource settings. Ethiop J Health Sci 2016 May;26(3):277-284. [doi: 10.4314/ejhs.v26i3.10] [Medline: 27358548]
59. Campos VM, Maggitti-Bezerril M, Menezes RC, Sobral L, Avena KM, Andrade BB. Doctors in the making, or degrees for sale? A student's view of Brazil's medical education crisis. Front Med (Lausanne) 2025;12:1687117. [doi: 10.3389/fmed.2025.1687117] [Medline: 41189902]
60. Andrade BB. The dark side of private medical education in Brazil. Front Med (Lausanne) 2025;12:1504794. [doi: 10.3389/fmed.2025.1504794] [Medline: 40034387]
61. Andrade BB, Villalva-Serra K, Menezes RC, Quintanilha LF, Avena KM. For-profit growth and academic decline: a retrospective nationwide assessment of Brazilian medical schools. Front Med (Lausanne) 2025;12:1617885. [doi: 10.3389/fmed.2025.1617885] [Medline: 40687715]




XSL•FO
RenderX

62. Charoenporn V, Hanvivattanakul S, Jongmekwamsuk K, Lenavat R, Hanvivattanakul K, Charernboon T. Zoom fatigue related to online learning among medical students in Thailand: prevalence, predictors, and association with depression. F1000Res 2024;13:617. [doi: 10.12688/f1000research.146084.2] [Medline: 39220383]
63. Ibrahim RK, Al Marar YA, Salman M, et al. Impact of multiple educational technologies on well-being: the mediating role of digital cognitive load. BMC Nurs 2025 Aug 5;24(1):1028. [doi: 10.1186/s12912-025-03655-z] [Medline: 40764954]
64. Izquierdo-Condoy JS, Arias-Intriago M, Montero Corrales L, Ortiz-Prado E. Artificial intelligence in medical education: transformative potential, current applications, and future implications. JMIR Med Educ 2026 Feb 17;12:e77127. [doi: 10.2196/77127] [Medline: 41701936]
65. Dias AK, De Melo TS, Avena KD, Andrade BD, Quintanilha LF. Artificial intelligence in health education: teachers' and tech experts' views on replacing teaching competencies. Int J Educ Health 2025;9:e6382. [doi: 10.17267/2594-7907ijeh.2025.e6382]
66. Dedivitis RA, Matos LL, Castro MA, Castro AA, Giaxa RR, Tempski PZ. Empathy and spirituality in medical students and residents [Article in Portuguese]. Rev Bras Educ Med 2024;48(4):e104. [doi: 10.1590/1981-5271v48.4-2023-0042]
67. Hamilton A, Molzahn A, McLemore K. The evolution from standardized to virtual patients in medical education. Cureus 2024 Oct;16(10):e71224. [doi: 10.7759/cureus.71224] [Medline: 39525234]
68. Kay D, Pasarica M. Using technology to increase student (and faculty satisfaction with) engagement in medical education. Adv Physiol Educ 2019 Sep 1;43(3):408-413. [doi: 10.1152/advan.00033.2019] [Medline: 31408381]
69. Oliveira NM, Viana DA, Santos JR, et al. Engagement in extracurricular activities during medical school: a cross-sectional study on student motivations and challenges. J Med Educ Curric Dev 2024;11:23821205241296980. [doi: 10.1177/23821205241296980] [Medline: 39600962]
70. Pelaccia T, Viau R. Motivation in medical education. Med Teach 2017 Feb;39(2):136-140. [doi: 10.1080/0142159X.2016.1248924] [Medline: 27866457]
71. Gruppen LD, Mangrulkar RS, Kolars JC. The promise of competency-based education in the health professions for improving global health. Hum Resour Health 2012 Nov 16;10:43. [doi: 10.1186/1478-4491-10-43] [Medline: 23157696]


**Abbreviations**

**LMIC:** low- and middle-income country
**PBL:** problem-based learning
**STEAM:** science, technology, engineering, arts, and mathematics


*Edited by M Balcarras; submitted 31.Mar.2026; peer-reviewed by X Wu; revised version received 29.Jun.2026; accepted 02.Jul.2026; published 28.Jul.2026.*

*Please cite as:*
*Avena KM, Brito APA, Quintanilha LF, Araújo-Pereira M, Andrade BB*
*Rethinking Basic Science Teaching in Medical Education: A Viewpoint on Its Past, Present, and Future*
*Interact J Med Res 2026;15:e80646*
*URL: https://www.i-jmr.org/2026/1/e80646*
*doi:10.2196/80646*


© Kátia M Avena, Ana Paula A Brito, Luiz F Quintanilha, Mariana Araújo-Pereira, Bruno B Andrade. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 28.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Health Care Analytics Challenges: A 3-Pillar Framework Connecting Analytics Maturity, Workforce Agility, and Technical Enablement

Samuel Thomas Harrold

Yuimedi, Inc., 240 Elm Street, 2nd Floor, Somerville, MA, United States

**Corresponding Author:**
Samuel Thomas Harrold
Yuimedi, Inc., 240 Elm Street, 2nd Floor, Somerville, MA, United States

## Abstract

Health care organizations face a "triple threat" of low analytics maturity, high workforce instability, and semantic technical barriers that together produce a crisis of "institutional amnesia." Leadership turnover, workforce shortages, and widespread intent to leave among informatics specialists systematically erase the tacit knowledge required to navigate complex clinical data schemas, trapping organizations in a cycle where knowledge loss outpaces knowledge capture. Viewed through the socialization, externalization, combination, and internalization model of knowledge creation by Nonaka, the root cause is a "socialization failure": high turnover fractures the social networks required for mentorship, rendering the traditional apprenticeship model of informatics unsustainable. To address this failure, we used a design science research approach synthesizing evidence from health care informatics, knowledge management, and natural language processing to develop a sociotechnical framework: human-in-the-loop knowledge governance (HITL-KG). HITL-KG is designed to shift the locus of organizational knowledge from volatile human memory to durable semantic artifacts called "validated query triples," each comprising a natural language intent, executable SQL, and rationale metadata. By embedding knowledge capture into the daily query workflow, the framework aims to convert ephemeral analytics into permanent institutional assets. The accompanying 3-pillar assessment rubric enables organizations to identify compounding vulnerabilities across analytics maturity, workforce agility, and technical enablement. The "validator paradox" (who validates the AI when experts leave?) is addressed by reframing validation through lean "standard work": each validated query establishes the current known standard rather than eternal truth, functioning as a "knowledge ratchet" that prevents regression. Decoupling analytical capability from individual tenure lets analytics maturity advance even as the workforce evolves. This paper proposes and theoretically motivates the framework; empirical validation is deferred to a companion study.

***(Interact J Med Res 2026;15:e96541)*** doi:10.2196/96541

**KEYWORDS**

institutional amnesia; organizational forgetting; human in the loop; medical informatics; sociotechnical systems; query governance; natural language processing; knowledge management; personnel turnover; organizational resilience; artificial intelligence; AI

## The Triple Threat: Institutional Amnesia in Health Care Analytics

The health care analytics landscape is currently paralyzed by a "triple threat" of compounding failures: (1) persistently *low analytics maturity*, where despite decades of investment, as of late 2024, only 39 organizations worldwide had reached the top tiers of the Healthcare Information and Management Systems Society (HIMSS) Analytics Maturity Assessment Model (AMAM) [1,2]; (2) a *semantic gap* between clinical intent and technical schema implementation [3,4]; and (3) a profound crisis of *workforce instability* that creates "institutional amnesia" [5].

While technical barriers and maturity models are well documented, the workforce dimension has shifted from a management concern to an existential threat. Modern longitudinal data on analytics staff are fragmented, but the available signals are alarming. As of 2024, a total of 53% of health care chief information officers have held their roles for less than 3 years [6], creating a strategic vacuum at the top. At the operational level, global nurse turnover runs at 16% to 18% annually [7,8], and a 2025 study found that 55% of public health informatics specialists intend to leave their positions [9].

This turnover creates what public administration scholars term "institutional amnesia" [10,11] and organization science studies refer to as "organizational forgetting" [12,13]: the systematic erasure of the tacit knowledge required to interpret complex health data in health care analytics. In health care, "data" are never raw; they are wrapped in layers of institutional context (billing rules, workflow work-arounds, and unwritten exclusions) [14]. When the analyst who knows that "exclusion code 99" actually means "hospice transfer" leaves, that



XSL•FO
RenderX

knowledge evaporates. The organization does not just lose an employee; it loses the ability to accurately measure its own performance.

Current literature approaches these problems in isolation. Analytics maturity models (eg, the HIMSS AMAM) assume a stable workforce capable of linear progression [1]. Technical solutions (eg, natural language to SQL; NL2SQL) assume a stable schema and clear intent [15]. Neither accounts for the reality of the "great resignation," where the rate of knowledge loss often exceeds the rate of knowledge capture [13].

Traditional knowledge management strategies (wikis, data dictionaries, and documentation) have failed because they are *passive* [16]. They require overworked staff to stop working and write down what they know. In a high-burnout environment, this documentation is the first casualty. As a result, health care systems are trapped in a Sisyphus-like cycle: hiring new analysts who spend their limited tenure relearning the same institutional secrets only to leave just as they become productive [17,18]. A foundational 2004 study established that technical IT staff in organizations where IT is a support function, as in health care delivery, had the lowest expected tenure of any organizational category at just 2.9 years [19]. That this 2-decade–old study remains a key benchmark is itself evidence of the crisis: the industry lacks the stability to track its own attrition. Contemporary signals, detailed below, suggest that the situation has worsened.

This viewpoint addresses a critical sociotechnical gap: *how can health systems maintain analytics maturity when workforce turnover exceeds the speed of documentation?*

As a viewpoint, this paper deliberately advances a prescriptive position: that health care organizations should shift from passive knowledge management to active, artifact-based governance. The analysis below is grounded in descriptive evidence of why current approaches fail, but the architectural recommendations are intentionally directive.

We propose that the solution lies not in better documentation but in a fundamental architectural shift: moving from *passive* knowledge management to *human-in-the-loop knowledge governance* (HITL-KG).

## Theoretical Grounding: Socialization, Externalization, Combination, and Internalization and the Unstable Workforce

### Overview

We ground our analysis in the socialization, externalization, combination, and internalization model of knowledge creation by Nonaka [20], informed by a narrative literature review across health care analytics maturity, workforce turnover, and natural language processing, with gray literature assessed using the authority, accuracy, coverage, objectivity, date, and significance checklist [21]. Gray literature was retained only when no peer-reviewed equivalent existed or when it provided unique industry data. The model describes organizational knowledge as emerging through a continuous cycle of 4 conversion modes:

1. *Socialization* (tacit to tacit)—knowledge transfers through shared experience and co-located practice, as when a senior analyst teaches a junior colleague the unwritten rules of a clinical dataset.
2. *Externalization* (tacit to explicit)—individuals articulate tacit know-how into explicit forms, such as documents or coded artifacts, as when an analyst records why a specific exclusion code maps to hospice transfers.
3. *Combination* (explicit to explicit)—separately documented knowledge is integrated and systematized into broader structures, as when data dictionary entries are consolidated into a governed analytics catalog.
4. *Internalization* (explicit to tacit)—individuals learn from documented knowledge and convert it into personal expertise through practice, as when a new analyst studies validated query libraries.

In a healthy organization, these 4 modes form a self-reinforcing spiral: tacit insights become documented, documentation becomes systematized, and systematized knowledge is internalized by new members who then generate fresh tacit insights [20]. When any mode breaks down, the spiral stalls. In health care analytics, the breakdown is at the very first stage.

The framework's 3 pillars name the organizational capabilities at stake in the "triple threat": analytics maturity, workforce agility, and technical enablement. This structure aligns with established models across health care informatics and knowledge management (Table 1).

**Table .** Framework alignment with established models.

| Pillar | HIMSS[a] AMAM[b] alignment | DIKW[c] hierarchy | Knowledge management |
|---|---|---|---|
| Analytics maturity | Progression from stages 0-7 | Data→information | Organizational learning |
| Workforce agility | Implicit in advanced stages | Knowledge (tacit)→wisdom | Tacit knowledge transfer |
| Technical enablement | Stages 6-7 requirements | Information→knowledge | Knowledge codification |

[a]HIMSS: Healthcare Information and Management Systems Society.

[b]AMAM: Analytics Maturity Assessment Model.

[c]DIKW: data, information, knowledge, and wisdom.

The HIMSS AMAM provides organizational benchmarks but does not address workforce knowledge retention. The data, information, knowledge, and wisdom hierarchy explains progression from raw data to actionable insights but does not



XSL•FO
RenderX

account for institutional memory loss. The 3-pillar framework synthesizes these perspectives, positioning workforce dynamics as the critical enabler connecting data access (analytics maturity) with organizational wisdom (knowledge preservation) [13,20].

### The Broken Cycle: Socialization Failure

In the model by Nonaka, *socialization* is the foundational conversion mode: the primary channel through which newcomers absorb tacit context that formal training cannot convey [20]. Socialization depends on 2 preconditions: sustained interaction and sufficient temporal overlap between knowledge holders and receivers [22]. It is, in effect, an apprenticeship model requiring years of shared practice.

In the current health care environment, this mechanism has collapsed. The leadership and operational turnover documented above reveals an "apprenticeship window" shorter than the knowledge transfer cycle it must support: 30% of new employees depart within their first year [23], whereas specialized informatics roles require 18 to 24 months to reach fluency [17]. The arithmetic is unforgiving: by the time a new analyst has absorbed enough tacit context to be productive, their mentor may already be gone, and the new analyst is themselves halfway through an average tenure.

High turnover rates fracture the social networks required for mentorship [7,8]. The resulting knowledge loss is compounding: each departure removes a node from the organization's informal knowledge network, making subsequent socialization attempts less effective because fewer experienced practitioners remain to serve as mentors [24]. Socialization is no longer a viable strategy for resilience.

### The Solution: Externalization via Sociotechnical Artifacts

To survive, organizations must shift reliance from socialization to *externalization*: converting tacit knowledge into explicit, durable artifacts [25]. However, traditional externalization (writing wikis, maintaining data dictionaries, and composing runbooks) suffers from 2 critical weaknesses. First, it is passive: it requires overworked staff to interrupt their workflow and perform a separate documentation task [26]. In a high-burnout environment where organizations face persistent attrition and talent shortages [27], this discretionary documentation is the first casualty. Second, it is low fidelity: the act of writing down tacit knowledge inevitably loses nuance, context, and the conditional logic that makes institutional knowledge valuable [22]. The result is documentation that exists but does not adequately capture what the departing expert actually knew.

We propose a form of *active* externalization: one that captures tacit knowledge as a by-product of the daily analytics workflow rather than a separate documentation burden. The mechanism is a new sociotechnical artifact: the *validated query triple* (see Multimedia Appendix 1 for worked examples). This artifact consists of (1) *natural language intent* (the clinical business question, eg, "Hypertension readmissions excluding planned transfers"), (2) *executable SQL* (the technical implementation), and (3) *rationale metadata* (the "why" behind the logic, eg, "Excluding status 02 per Centers for Medicare and Medicaid Services (CMS) 2025 rule").

By capturing these 3 components *during the act of analytics*, we transform the ephemeral work of query generation into a permanent institutional asset [28].

## HITL-KG

We propose HITL-KG as the overarching governance framework, extending the knowledge governance approach [29] with the validated query cycle as its core operational process. This framing reflects that the system serves as a governance mechanism, not just a productivity tool.

### The HITL-KG Architecture

The HITL-KG architecture (Figure 1) functions as a *governance forcing function*. It inserts a mandatory validation step into the analytics workflow, preventing laundered hallucinations while capturing expert knowledge.

A clinical user's natural language query (step 1) is translated to SQL and run against the data warehouse (steps 2-3); an expert validation gate (step 4) either returns confirmed insights (steps 5-6) or loops back for correction. Confirmed query triples enter organizational memory (step 7; dashed line in Figure 1), which informs future queries and curates the knowledge base (step 8), closing a continuous learning loop in which best practices evolve rather than remain static.



XSL•FO
RenderX

**Figure 1.** Human-in-the-loop knowledge governance architecture. NLP: natural language processing.

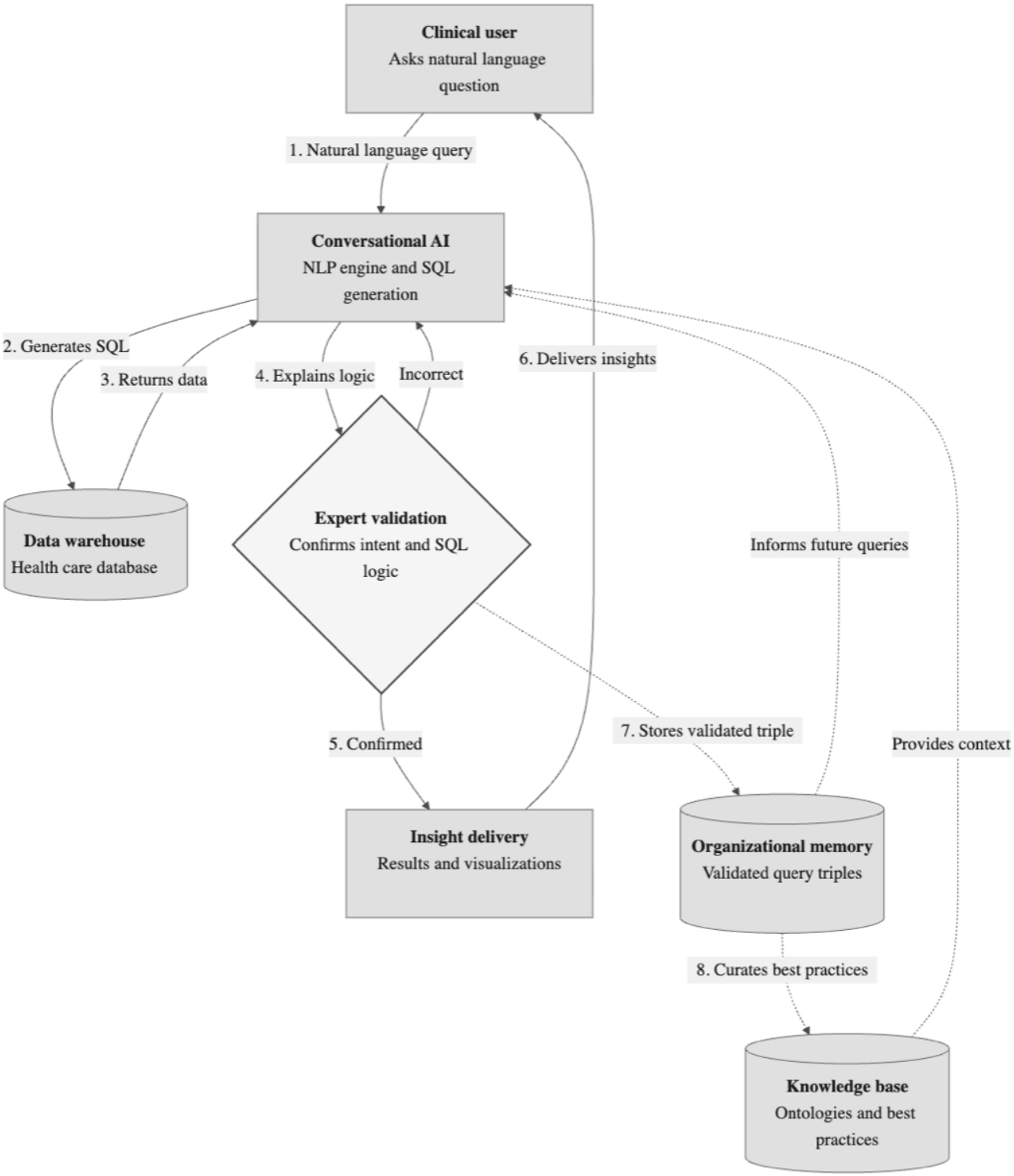



Figure 2 summarizes the corresponding 6-step validated query cycle, tracing queries from clinical intent to expert validation into durable organizational memory.

A natural language question (step 1) yields a candidate SQL (step 2) that the analyst validates against the system's explained logic (step 3), looping back if incorrect. Validated triples are stored in organizational memory (step 4) matched by future queries (step 5) and persist across staff turnover (step 6). The dashed arrows in Figure 2 mark the pillar outcomes: storage advances analytics maturity, persistence stabilizes workforce agility, and retrieval increases technical enablement.



XSL•FO
RenderX

**Figure 2.** The 6-step validated query cycle and its mapping to the 3 pillars.



XSL•FO
RenderX

### The Process of Externalization

The cycle proceeds from query generation, in which a conversational AI system translates the user's natural language question into candidate SQL [30], to *expert validation*, the critical moment of externalization where the domain expert confirms or corrects the AI's interpretation and thereby captures tacit knowledge [31,32]. The validated triple is hashed and stored in organizational memory [33], then retrieved ahead of future probabilistic generation [34]. Multimedia Appendix 1 details each step alongside worked examples.

### Comparison With Existing Approaches

Organizations have addressed institutional memory loss through several strategies, each with limitations that HITL-KG overcomes.

*Code-based semantic layers* (eg, data build tool and LookML) encode business logic in version-controlled repositories but suffer from "schema rot," where electronic medical record (EMR) data models change frequently and maintenance exceeds high-turnover teams' capacity, misaligning the layer and the underlying data [35]. HITL-KG's validated query triples share the version control principle but add the rationale metadata that semantic layers lack: not just *what* a query does but *why*.

*Traditional knowledge management* (wikis, data dictionaries, and runbooks) relies on passive capture, where users must stop working to document, which reduces participation and produces inaccurate records under cognitive load [16,26]. Such documentation might suffice for a small, stable query set, but real workloads are neither: more than half of analytical queries recur in large production workloads [36], yet even small databases log hundreds to thousands of distinct query strings [37], so a fixed set of documented queries captures neither the large recurring core nor the evolving ad hoc tail. HITL-KG instead implements active capture across both regimes: the query itself is the documentation, validated at the point of use [28].

*Unsupervised AI querying* removes human oversight entirely; current NL2SQL accuracy makes this unsafe for clinical analytics [38]. HITL-KG occupies the middle ground: AI generates, humans validate.

## The Evidence Base: 3 Pillars

### Overview

The evidence presented below supports the existence and severity of the problem that HITL-KG addresses and the maturity of the technologies it requires. It does not constitute empirical validation of the framework itself, which remains a theoretically grounded, testable proposition whose empirical evaluation is the subject of future work.

The HITL-KG framework is supported by 3 pillars of empirical evidence synthesized from the literature cited throughout this paper.

### Pillar 1: Analytics Maturity Evidence

*Analytics maturity* describes an organization's progression in using data and quantitative models for fact-based decisions [39], a construct operationalized by staged maturity models that extend beyond the HIMSS AMAM [40,41]. Health care maturity remains chronically low. Assessments reveal that only 26 organizations worldwide achieved AMAM stage 6 and 13 reached stage 7 by late 2024 [1,2]. Because AMAM assessment is voluntary and self-selected, these figures likely overrepresent analytics-committed organizations, so low maturity may be even more widespread than the counts suggest [42,43]. The corollary is that the vast majority of organizations remain at the lower stages, characterized by fragmented data and limited predictive capabilities [44]. However, maturity is not merely an IT metric; it is a clinical safety predictor. In HIMSS's companion, the EMR adoption model, which measures the EMR adoption underlying analytics capability, levels 6 and 7 correlate with 3.25 times higher odds of better Leapfrog Hospital Safety Grades [45]. Low maturity creates a "low maturity trap" where data quality issues (such as the 39%-71% missing data rates in cancer databases [46]) remain uncorrected because the experts who understand the context are leaving.

Critically, low maturity is not simply a status; it is a self-reinforcing trap. Organizations at stages 0 to 3 lack the automated monitoring and data governance infrastructure needed to detect their own deficiencies, so poor data quality goes unrecognized. The clinical consequences are measurable: one Medicare accountable care organization that implemented analytics to overcome electronic health record fragmentation reduced readmission rates from 24% to 17.8% and achieved US $1.6 million in cost savings [47]. However, data interoperability remains a leading obstacle to analytics adoption [3,4]. Barriers including employee resistance to change and lack of organizational readiness further stall data-driven initiatives [48,49]. The trap thus compounds: each year of delayed investment widens the gap between what the organization could know and what it does know.

### Pillar 2: Workforce Agility Evidence

*Workforce agility* is a workforce's capacity for proactivity, adaptability, and resilience under change [50,51]. In health care analytics, the evidence below documents the instability that systematically erodes this capacity. The cost of turnover in informatics is higher than in standard IT. Knowledge loss can cost up to 3 times the annual salary [24,52]. In clinical informatics specifically, replacement costs range from US $50,000 to US $230,000 per departing employee, with a loss of 39 staff members in a single department in 1 year incurring a total cost of up to US $8.97 million [53]. With 30% of new employees leaving within their first year [23], health care IT professionals spend a limited portion of their employment at full productivity as specialized roles require 18 to 24 months to reach fluency [17]. This "revolving door" prevents the accumulation of the "collective knowledge structures" required for complex task performance [13].

The workforce crisis operates at multiple reinforcing levels, as detailed in the Theoretical Grounding: Socialization, Externalization, Combination, and Internalization and the Unstable Workforce section: leadership churn, operational shortages, and foundational attrition collectively prevent the sustained mentorship that socialization requires. This multilevel



XSL•FO
RenderX

instability ensures that the temporal overlap between experienced practitioners and newcomers rarely materializes.

### Pillar 3: Technical Enablement Evidence

*Technical enablement* denotes the technologies and practices that let nonspecialist domain users access and analyze data without heavy reliance on central IT [54,55]. NL2SQL technology has matured along a clear accuracy gradient: general-purpose models achieve approximately 65% execution accuracy on health care–specific benchmarks such as MIMICSQL [56], domain-adapted systems such as MedT5SQL reach 80% accuracy [57], and architecturally specialized approaches combining large language models (LLMs) with structured knowledge representations achieve 94% accuracy on the same benchmarks [58]. While insufficient for unsupervised clinical deployment [38], this gradient demonstrates that the generation engine required for HITL-KG is viable within a human-validated workflow. Notably, high benchmark accuracy does not guarantee real-world generalization: evaluation on more challenging dataset splits reveals accuracy drops from 92% to 28% [59], reinforcing the necessity of the human-in-the-loop architecture. Health care–specific natural language interfaces such as Criteria2Query achieve fully automated cohort query formulation in 1.22 seconds per criterion [60], with over 80% of clinical users indicating willingness to adopt such tools.

However, NL2SQL is an enabler, not a solution in itself. Its deeper significance lies in the organizational prerequisites it demands. For an AI system to translate a natural language question into a correct SQL query, the organization must first establish validated data models, explicit business logic definitions, and codified domain terminology [3,4]. This reframes NL2SQL from a convenience tool into a catalyst for the kind of systematic knowledge externalization that the HITL-KG framework requires. The interface bridges the semantic gap between clinical experts and technical schema, allowing nontechnical domain experts to interact with data alongside broader modernization efforts [61-64]. The very difficulty of making NL2SQL work correctly thus becomes a governance opportunity: it compels organizations to surface and formalize tacit knowledge that would otherwise depart with its experts.

## Organizational Self-Assessment

### Overview

To operationalize the framework, we propose a *3-pillar assessment rubric* (Table 2) that enables health care organizations to evaluate their current position across each pillar and identify compounding vulnerabilities.

**Table .** Three-pillar assessment rubric indicators (the full rubric with anchors can be found in Multimedia Appendix 2).

| Pillar | Indicators |
|---|---|
| Analytics maturity | HIMSS[a] AMAM[b] stage, self-service analytics, and AI and NL[c] interface |
| Workforce agility | First-year analytics staff turnover, leadership tenure, and knowledge concentration |
| Technical enablement | Data access, interoperability, and schema coupling |

[a]HIMSS: Healthcare Information and Management Systems Society.

[b]AMAM: Analytics Maturity Assessment Model.

[c]NL: natural language.

### Why Assessment, Not Just Maturity

As noted, the maturity models and knowledge management frameworks in Table 1 assume linear progression by a stable workforce [1]. In practice, an organization that reaches AMAM stage 5 but regresses to stage 3 after the departure of 2 senior analysts has not failed to mature; it has failed to be *resilient*. A recent systematic review of organizational resilience measurement in health care found no consensus on what to measure [65]. Only 4 instruments have been developed specifically for health care, and only 2 have been validated [66]. The 3-pillar assessment rubric addresses this gap by organizing evidence-based indicators around the domains where institutional amnesia operates, measuring not just where an organization stands but also how vulnerable it is.

### The 3-Pillar Rubric

Each indicator is scored as being of low, medium, or high strength using evidence-based anchors from the literature reviewed in the The Evidence Base: 3 Pillars section. Organizations scoring predominantly as “low strength” across multiple pillars face the self-reinforcing degradation cycle that the framework identifies as the central threat. Table 2 summarizes the 9 indicators; the full rubric with scoring anchors and evidence can be found in Multimedia Appendix 2.

The “schema coupling” indicator can be operationalized through *continuous analytic integration*: treating validated query triples as software assets within a continuous integration and continuous delivery pipeline that detects data and schema drift [35,67,68]. When a data warehouse schema is updated, the system automatically reruns stored queries and flags failures, transforming institutional memory into a living test suite [69,70].

The rubric complements rather than replaces the AMAM. Where the AMAM measures the sophistication of analytical capabilities at a point in time, the 3-pillar assessment rubric reveals how vulnerable those capabilities are to the compounding effects of turnover, low maturity, and technical barriers. Used together, they provide a 2D view of analytics health.



XSL•FO
RenderX

## The Validator Paradox and Standard Work

A critical objection to HITL-KG is circular: if the framework requires domain experts to validate AI-generated queries and the core problem is that domain experts are leaving, then the framework fails precisely when it is most needed. This *validator paradox* represents the strongest counterargument to the approach proposed here, and addressing it requires moving beyond simplistic reassurance.

The resolution draws on lean management's concept of "standard work" [71]. In this framing, validation is not the establishment of *eternal truth* but the documentation of the *current known standard*. Each validated query triple records the best available understanding of how a business question maps to a data operation at that moment; the validation is time-stamped and contextual, not permanent. Critically, as Alukal and Manos [71] establish, standard work is the prerequisite for kaizen (continuous improvement): without a documented baseline, there is no foundation to improve upon. Each validated query, therefore, establishes a floor, not a ceiling. The next expert, veteran or midcareer hire, inherits a baseline to refine rather than reconstructing institutional knowledge from scratch.

This mechanism functions as what we term a *knowledge ratchet*, consistent with findings on collective knowledge structures [13]. Each validated triple prevents regression below the last confirmed state. Even if a subsequent validator is less experienced, the organization cannot slide below the previously validated standard. The ratchet does not guarantee forward progress, but it prevents the catastrophic backsliding that is the central failure mode of institutional amnesia.

Real-world evidence supports this: UC Davis Health moved from AMAM stage 0 to stage 6 by establishing standardized "S.M.A.R.T." definitions for its analytics metrics [72]. Those codified standards survived staff turnover because they existed as organizational artifacts rather than knowledge held solely by their creators. The HITL-KG validated query library serves an analogous function: it encodes analytical decisions into durable, retrievable structures that persist independent of any single analyst's tenure.

The validator paradox is not fully resolved: below a minimum viable expertise threshold, validation becomes meaningless. A junior analyst rubber-stamping AI-generated output without genuine comprehension provides no knowledge ratchet; the validation artifact exists, but its epistemic value is nil. Identifying this threshold remains an open empirical question; future work should measure it via controlled hallucination injection studies in which AI-generated queries containing deliberate errors are presented to validators of varying experience levels.

We propose a provisional 3-tier governance model informed by risk-stratified AI oversight frameworks [73,74] to degrade gracefully when expertise is scarce. The first tier is *full validation*: a domain expert with schema-level knowledge reviews the query triple and confirms semantic alignment (the default). The second tier is *constrained validation*: when no fully qualified validator is available, a less experienced analyst reviews the triple against previously validated queries for the same data domain, flagging deviations for deferred expert review; the triple is stored with "provisional" status. The third tier is *automated matching*: for queries that match an existing validated triple with high semantic similarity, the system accepts the match without human review, logging it for periodic audit.

Organizational governance requirements include who can validate queries (domain expertise thresholds), review workflows for high-stakes queries, query versioning as schemas evolve, and retrieval policies. Drawing on established master data governance practice [75], organizations can designate specific validated triples as "golden queries" certified by a governance committee as the authoritative source of truth for key metrics. Only these certified triples serve as official standards, mitigating shadow IT risks [76] while preserving analytical agility.

## Safety as Cognitive Forcing

HITL-KG is fundamentally a *safety mechanism*, not a productivity tool. The central risk of unsupervised AI in clinical analytics is what we term "laundering hallucinations": a plausible-sounding but factually incorrect query result that enters the decision pipeline undetected and influences clinical or operational choices. Because LLMs generate fluent, confident output regardless of correctness, errors arrive dressed in the language of expertise, harder to catch than obviously malformed output.

HITL-KG mitigates this risk through *cognitive forcing functions* [38], a design pattern borrowed from clinical decision-making and safety engineering. Requiring the AI to explain its reasoning *before* presenting results forces the validator into system 2 (analytical and deliberate) thinking rather than system 1 (fast and heuristic) acceptance of superficially plausible outputs. User studies confirm the practical benefit: structured explanation reduces error recovery time by 30 to 40 seconds compared to unstructured output review [77].

Critically, the human validator's value is not merely assumed. Recent studies show that human oversight measurably improves AI-generated analytics: LLM output verified by a reviewer outperformed a human-only workflow (91.0% vs 89.0% accuracy) across 6 systematic reviews [78], and a text-to-SQL validation gate in which experts corrected generated queries improved semantic correctness by 28.4% over end-to-end model output [79]. These gains are conditional on well-designed interaction [80]; the comparative effectiveness of HITL-KG itself remains the subject of the planned empirical validation.

The parallel to aviation safety is instructive: checklists and mandatory callouts made commercial aviation exceptionally safe not by eliminating human error but by surfacing errors before they propagate. HITL-KG applies the same principle: the mandatory validation step is an analytical "callout" that interrupts uncritical acceptance so that the friction of validation is not a cost but the mechanism of safety itself.



XSL•FO
RenderX

## Structural Barriers: Why the Problem Persists

Failed standardization approaches (eg, IBM Watson Health [81,82] and Haven [83,84]) demonstrate that centralized models fail clinical reality. Metadata uncertainties and "messy" institution-specific business logic require localized solutions [3,82]. HITL-KG addresses this by capturing *local* logic rather than enforcing *global* standards.

## Limitations

This work is a narrative, design science–informed framework rather than a systematic review or multisite empirical evaluation. The literature base is concentrated on English-language sources and recent (2024-2026) workforce and NL2SQL studies, so findings may not capture all regional, specialty-specific, or technological contexts. The HITL-KG architecture and the proposed 3-pillar assessment rubric are conceptual artifacts that require future implementation and validation in diverse health systems before their effectiveness and generalizability can be fully established.

HITL-KG also complements rather than replaces workforce development: it decouples institutional knowledge from individual tenure but still presumes investment in the training and analytics literacy through which staff learn to pose, validate, and interpret queries [17]. This complementarity also bounds the framework. Where data models and business rules change continually, no approach is self-maintaining; precisely because retained expertise cannot manually keep pace with constant change, knowledge must be externalized into curated artifacts, yet the validated query library must itself be actively maintained, a burden that the continuous analytic integration mechanism mitigates but does not remove.

## Implications and Future Research

The crisis of institutional amnesia in health care requires a structural shift. As long as analytical maturity is tied to individual tenure, organizations will remain fragile. By implementing HITL-KG, health systems can decouple intelligence from turnover, building a library of validated knowledge that ensures that maturity advances even as the workforce evolves.

Future research should empirically validate and refine the HITL-KG framework and the proposed 3-pillar assessment rubric. Priority questions include how pillar scores correlate with observed continuity of analytics performance during leadership and staff turnover; whether HITL-KG–mediated NL2SQL workflows reduce error rates, recovery time, and rework compared to baseline tooling; and which governance patterns most effectively balance safety, transparency, and equity when human validators operate at scale. Prospective multisite implementation studies, controlled user experiments, and qualitative implementation research will be needed to test these claims across diverse organizational, regulatory, and data environments.

### Acknowledgments

STH is the sole author and takes full responsibility for the manuscript's content. Generative AI was used to assist this work, specifically Gemini CLI (Gemini 3; Google) and Claude Code (Claude 5; Anthropic) for manuscript editing, language refinement, and literature search. Generative AI was not used to generate research findings, data, or conclusions; the author conducted the research and verified all claims, citations, and AI-assisted text. Figures were generated using the Mermaid graph language.

### Funding

Yuimedi, Inc., provided financial support for the author's time researching and writing this manuscript. The views expressed are the author's own and do not represent the official positions of Indiana University Health or Yuimedi, Inc.

### Data Availability

Data sharing is not applicable to this article as no data sets were generated or analyzed during this study. All data cited are from publicly available peer-reviewed publications, industry reports, and academic theses referenced in the bibliography. The manuscript source, build pipeline, and literature review code generated for this work are archived on Zenodo [85].

### Authors' Contributions

Conceptualization: STH
Investigation: STH
Methodology: STH
Visualization: STH
Writing—original draft: STH
Writing—review and editing: STH

### Conflicts of Interest

STH is a contract product advisor at Yuimedi, Inc., which develops health care analytics software, including conversational AI platforms relevant to this review's subject matter. The author is also employed as a data scientist at Indiana University Health.



XSL•FO
RenderX

This paper presents an analytical framework derived from published literature and does not evaluate or recommend specific commercial products, including those of the author's affiliated organizations.

Multimedia Appendix 1
Validated query triple examples illustrating the human-in-the-loop knowledge governance knowledge capture artifact. Each triple comprises a natural language intent, executable SQL, and rationale metadata for 3 clinical analytics scenarios: diabetes care monitoring, heart failure readmission cohort identification, and quality metric aggregation.
[DOCX File, 15 KB - i-jmr_v15i1e96541_app1.docx ]

Multimedia Appendix 2
Three-pillar organizational assessment rubric: 9 indicators across analytics maturity, workforce agility, and technical enablement, each with low, medium, and high strength scoring anchors and supporting evidence from the literature.
[DOCX File, 16 KB - i-jmr_v15i1e96541_app2.docx ]

## References


1. Analytics Maturity Assessment Model (AMAM). Healthcare Information and Management Systems Society. URL: https://www.himss.org/maturity-models/amam/ [accessed 2026-08-07]
2. At HIMSS24 APAC, the Adoption Model for Analytics Maturity gets facelift. Healthcare IT News. 2024. URL: https://www.healthcareitnews.com/news/asia/himss24-apac-adoption-model-analytics-maturity-gets-facelift [accessed 2026-08-07]
3. Gal MS, Rubinfeld DL. Data standardization. N Y Univ Law Rev 2019;94(4):737-770 [FREE Full text]
4. Zhang Y, Callaghan-Koru JA, Koru G. The challenges and opportunities of continuous data quality improvement for healthcare administration data. JAMIA Open 2024 Aug;7(3):ooae058. [doi: 10.1093/jamiaopen/ooae058] [Medline: 39091510]
5. Hong JH. When does employee turnover matter? Analyzing the role of organizational memory in the federal IT workforce. J Public Adm Res Theory 2025;35(4):434-451. [doi: 10.1093/jopart/muaf019]
6. CIO insights: the state of healthcare IT leadership. WittKieffer. 2024. URL: https://api.wittkieffer.com/wp-content/uploads/2012/10/cio-insights-the-state-of-healthcare-it-leadership-wittkieffer-october-2024.pdf [accessed 2026-08-07]
7. Ren H, Li P, Xue Y, Xin W, Yin X, Li H. Global prevalence of nurse turnover rates: a meta-analysis of 21 studies from 14 countries. J Nurs Manag 2024;2024:5063998. [doi: 10.1155/2024/5063998] [Medline: 40224762]
8. Wu F, Lao Y, Feng Y, Zhu J, Zhang Y, Li L. Worldwide prevalence and associated factors of nursing staff turnover: a systematic review and meta-analysis. Nurs Open 2024 Jan;11(1):e2097. [doi: 10.1002/nop2.2097] [Medline: 38268271]
9. Rajamani S, Leider JP, Gunashekar DR, Dixon BE. Public health informatics specialists in state and local public health workforce: insights from public health workforce interests and needs survey. J Am Med Inform Assoc 2025 Apr 1;32(4):748-754. [doi: 10.1093/jamia/ocaf019] [Medline: 39936849]
10. Pollitt C. Institutional amnesia: a paradox of the "information age"? Prometheus 2000;18(1):5-16. [doi: 10.1080/08109020050000627]
11. Stark A, Head B. Institutional amnesia and public policy. J Eur Public Policy 2019;26(10):1521-1539. [doi: 10.1080/13501763.2018.1535612]
12. de Holan PM, Phillips N. Remembrance of things past? The dynamics of organizational forgetting. Manag Sci 2004 Nov;50(11):1463-1613. [doi: 10.1287/mnsc.1040.0273]
13. Rao RD, Argote L. Organizational learning and forgetting: the effects of turnover and structure. Eur Manag Rev 2006;3(2):77-85. [doi: 10.1057/palgrave.emr.1500057]
14. Gitelman L, editor. "Raw Data" Is an Oxymoron: MIT Press; 2013. [doi: 10.7551/mitpress/9302.001.0001]
15. Wang P, Shi T, Reddy CK. Text-to-SQL generation for question answering on electronic medical records. In: The Web Conference 2020 - Proceedings of the World Wide Web Conference: Association for Computing Machinery; 2020. [doi: 10.1145/3366423.3380120]
16. Mayo CS, Deasy JO, Chera BS, Freymann J, Kirby JS, Hardenberg PH. How can we effect culture change toward data-driven medicine? Int J Radiat Oncol Biol Phys 2016 Jul 1;95(3):916-921. [doi: 10.1016/j.ijrobp.2015.12.355] [Medline: 27302507]
17. Ledikwe JH, Reason LL, Burnett SM, et al. Establishing a health information workforce: innovation for low- and middle-income countries. Hum Resour Health 2013 Jul 18;11:35. [doi: 10.1186/1478-4491-11-35] [Medline: 23866794]
18. Mantas J, Ammenwerth E, Demiris G, et al. Recommendations of the International Medical Informatics Association (IMIA) on education in biomedical and health informatics. First revision. Methods Inf Med 2010 Jan 7;49(2):105-120. [doi: 10.3414/ME5119] [Medline: 20054502]
19. Ang S, Slaughter S. Turnover of information technology professionals: the effects of internal labor market strategies. ACM SIGMIS DATABASE: the DATABASE for Advances in Information Systems 2004;35(3):11-27. [doi: 10.1145/1017114.1017118]
20. Farnese ML, Barbieri B, Chirumbolo A, Patriotta G. Managing knowledge in organizations: a Nonaka's SECI model operationalization. Front Psychol 2019;10:2730. [doi: 10.3389/fpsyg.2019.02730] [Medline: 31920792]




XSL•FO
RenderX

21. Tyndall J. AACODS checklist. Flinders University. 2010. URL: https://fac.flinders.edu.au/dspace/api/core/bitstreams/e94a96eb-0334-4300-8880-c836d4d9a676/content [accessed 2026-08-07]
22. Foos T, Schum G, Rothenberg S. Tacit knowledge transfer and the knowledge disconnect. J Knowl Manag 2006 Jan;10(1):6-18. [doi: 10.1108/13673270610650067]
23. 2026 NSI national health care retention & RN staffing report. : NSI Nursing Solutions; 2026 URL: https://www.nsinursingsolutions.com/documents/library/nsi_national_health_care_retention_report.pdf [accessed 2026-08-07]
24. Massingham PR. Measuring the impact of knowledge loss: a longitudinal study. J Knowl Manag 2018;22(4):721-758. [doi: 10.1108/JKM-08-2016-0338]
25. Zhang W, Zhang W, Daim T, Yalçın H. AI challenges conventional knowledge management: light the way for reframing SECI model and Ba theory. J Knowl Manag 2025 Apr 30;29(5):1618-1654. [doi: 10.1108/JKM-02-2024-0203]
26. Goffin K, Koners U. Tacit knowledge, lessons learnt, and new product development. J Prod Innov Manag 2011 Mar;28(2):300-318. [doi: 10.1111/j.1540-5885.2010.00798.x]
27. The future of workforce. Healthcare Information and Management Systems Society. URL: https://www.himss.org/resources/the-future-of-workforce/ [accessed 2026-08-07]
28. Moore D, Ge X, Sirkin D, Stenholm D, Ju W. ActiveNavigator: toward real-time knowledge capture and feedback in design workspaces. Int J Eng Educ 2018;34(2):723-733 [FREE Full text]
29. Foss NJ. The emerging knowledge governance approach: challenges and characteristics. Organization 2007 Jan;14(1):29-52. [doi: 10.1177/1350508407071859]
30. Lee G, Hwang H, Bae S, et al. EHRSQL: a practical text-to-SQL benchmark for electronic health records. Presented at: 36th Conference on Neural Information Processing Systems; Nov 28 to Dec 9, 2022 URL: https://proceedings.neurips.cc/paper_files/paper/2022/file/643e347250cf9289e5a2a6c1ed5ee42e-Paper-Datasets_and_Benchmarks.pdf [accessed 2026-08-09]
31. Bravo Rocca GJ. Human-on-the-loop continual learning: data, knowledge and agents for model adaptation [PhD thesis]. : Universitat Politècnica de Catalunya; 2025 URL: https://www.tdx.cat/handle/10803/695722#page=1 [accessed 2026-08-09]
32. Mosqueira-Rey E, Hernández-Pereira E, Alonso-Ríos D, Bobes-Bascarán J, Fernández-Leal Á. Human-in-the-loop machine learning: a state of the art. Artif Intell Rev 2023 Apr;56:3005-3054. [doi: 10.1007/s10462-022-10246-w]
33. Benbya H, Passiante G, Aissa Belbaly N. Corporate portal: a tool for knowledge management synchronization. Int J Inf Manag 2004 Jun;24(3):201-220. [doi: 10.1016/j.ijinfomgt.2003.12.012]
34. Whittaker S, Tucker S, Swampillai K, Laban R. Design and evaluation of systems to support interaction capture and retrieval. Pers Ubiquit Comput 2008;12(3):197-221. [doi: 10.1007/s00779-007-0146-3]
35. Battula SK. Adaptive data quality management for multi-cloud healthcare warehouses: FHIR-aware semantics and unsupervised thresholding. Int J Artif Intell Data Sci Mach Learn 2025;6(4):218-226. [doi: 10.63282/3050-9262.IJAIDSML-V6I4P130]
36. Jindal A, Patel H, Roy A, et al. Peregrine: workload optimization for cloud query engines. In: SoCC '19: Proceedings of the ACM Symposium on Cloud Computing: Association for Computing Machinery; 2019:416-427. [doi: 10.1145/3357223.3362726]
37. Kul G, Luong DT, Xie T, Chandola V, Kennedy O, Upadhyaya S. Similarity metrics for SQL query clustering. IEEE Trans Knowl Data Eng 2018;30(12):2408-2420. [doi: 10.1109/TKDE.2018.2831214]
38. Ziletti A, D'Ambrosi L. Retrieval augmented text-to-SQL generation for epidemiological question answering using electronic health records. In: Proceedings of the 6th Clinical Natural Language Processing Workshop: Association for Computational Linguistics; 2024:47-53. [doi: 10.18653/v1/2024.clinicalnlp-1.4]
39. Davenport TH, Harris JG. Competing on Analytics: The New Science of Winning: Harvard Business Press; 2007.
40. Grossman RL. A framework for evaluating the analytic maturity of an organization. Int J Inf Manag 2018 Feb;38(1):45-51. [doi: 10.1016/j.ijinfomgt.2017.08.005]
41. Langer B. Understanding data & analytics maturity: a systematic review of maturity model composition. Schmalenbach J Bus Res 2025 Jun;77:205-227. [doi: 10.1007/s41471-024-00205-2]
42. Halbesleben JR, Whitman MV. Evaluating survey quality in health services research: a decision framework for assessing nonresponse bias. Health Serv Res 2013 Jun;48(3):913-930. [doi: 10.1111/1475-6773.12002] [Medline: 23046097]
43. Tomaskovic-Devey D, Leiter J, Thompson S. Organizational survey nonresponse. Adm Sci Q 1994 Sep;39(3):439-457. [doi: 10.2307/2393298]
44. The Healthcare Analytics Adoption Model: a roadmap to analytic maturity. Health Catalyst. URL: https://www.healthcatalyst.com/learn/insights/healthcare-analytics-adoption-model-roadmap-analytic-maturity [accessed 2026-08-09]
45. Snowdon A, Hussein A, Danforth M, Wright A, Oakes R. Digital maturity as a predictor of quality and safety outcomes in US hospitals: cross-sectional observational study. J Med Internet Res 2024 Aug 6;26:e56316. [doi: 10.2196/56316] [Medline: 39106100]
46. Yang DX, Khera R, Miccio JA, et al. Prevalence of missing data in the National Cancer Database and association with overall survival. JAMA Netw Open 2021 Mar 1;4(3):e211793. [doi: 10.1001/jamanetworkopen.2021.1793] [Medline: 33755165]




XSL•FO
RenderX

47. Latrella M, Baldasare L. Improving patient outcomes while reducing readmissions with data analytics. Manage Healthc 2024;8(3):212-220. [doi: 10.69554/RITW9729]
48. Shahbaz M, Gao C, Zhai L, Shahzad F, Hu Y. Investigating the adoption of big data analytics in healthcare: the moderating role of resistance to change. J Big Data 2019;6:6. [doi: 10.1186/s40537-019-0170-y]
49. Kamble SS, Gunasekaran A, Goswami M, Manda J. A systematic perspective on the applications of big data analytics in healthcare management. Int J Healthc Manag 2019;12(3):226-240. [doi: 10.1080/20479700.2018.1531606]
50. Breu K, Hemingway CJ, Strathern M, Bridger D. Workforce agility: the new employee strategy for the knowledge economy. J Inf Technol 2002 Mar;17(1):21-31. [doi: 10.1080/02683960110132070]
51. Tessarini Junior G, Saltorato P. Workforce agility: a systematic literature review and a research agenda proposal. Innovar 2021;31(81):155-167. [doi: 10.15446/innovar.v31n81.95582]
52. Wynendaele H, Clays E, Peeters E, DeJonghe Y, Van Hecke A, Trybou J. Understanding turnover in healthcare and welfare sectors of high-income countries: an umbrella review. BMC Health Serv Res 2025 Jun 5;25(1):806. [doi: 10.1186/s12913-025-12966-5] [Medline: 40474167]
53. Hackney A. Onboarding new hires in the Clinical Informatics and Practice Support (CIPS) department with the Tiered Skills Acquisition Model (TSAM): a program evaluation [PhD thesis]. : Jacksonville University; 2024 URL: https://search.proquest.com/openview/9a3a9ef301c01ff0129d07f685b8643f/1 [accessed 2026-08-09]
54. Alpar P, Schulz M. Self-service business intelligence. Bus Inf Syst Eng 2016 Apr;58:151-155. [doi: 10.1007/s12599-016-0424-6]
55. Bani-Hani I, Tona O, Carlsson S. Patterns of resource integration in the self-service approach to business analytics. In: Proceedings of the Annual Hawaii International Conference on System Sciences: IEEE Computer Society; 2020:5359-5368. [doi: 10.24251/HICSS.2020.659]
56. Blašković L, Tanković N, Lorencin I, Baressi Šegota S. Robust clinical querying with local LLMs: lexical challenges in NL2SQL and retrieval-augmented QA on EHRs. Big Data Cogn Comput 2025;9(10):256. [doi: 10.3390/bdcc9100256]
57. Marshan A, Almutairi AN, Ioannou A, Bell D, Monaghan A, Arzoky M. MedT5SQL: a transformers-based large language model for text-to-SQL conversion in the healthcare domain. Front Big Data 2024;7:1371680. [doi: 10.3389/fdata.2024.1371680] [Medline: 38988646]
58. Chen Q, Peng J, Song B, Zhou Y, Ji R. Graph-empowered text-to-SQL generation on electronic medical records. Pattern Recognit 2026 Jan;169:111800. [doi: 10.1016/j.patcog.2025.111800]
59. Tarbell R, Choo KK, Dietrich G, Rios A. Towards understanding the generalization of medical text-to-SQL models and datasets. AMIA Annu Symp Proc 2024;2023:669-678. [Medline: 38222436]
60. Yuan C, Ryan PB, Ta C, et al. Criteria2Query: a natural language interface to clinical databases for cohort definition. J Am Med Inform Assoc 2019 Apr 1;26(4):294-305. [doi: 10.1093/jamia/ocy178] [Medline: 30753493]
61. The code modernization playbook: transforming legacy systems with AI. Anthropic. 2025. URL: https://resources.anthropic.com/code-modernization-playbook [accessed 2026-08-09]
62. Hendrix GG, Sacerdoti ED, Sagalowicz D, Slocum J. Developing a natural language interface to complex data. ACM Trans Database Syst 1978 Jun;3(2):105-147. [doi: 10.1145/320251.320253]
63. Ogunwole O, Onukwulu EC, Joel MO, Adaga EM, Ibeh AI. Modernizing legacy systems: a scalable approach to next-generation data architectures and seamless integration. Int J Multidiscip Res Growth Eval 2025;4(1):901-909. [doi: 10.54660/.IJMRGE.2023.4.1.901-909]
64. Arora A. Challenges of integrating artificial intelligence in legacy systems and potential solutions for seamless integration. SSRN. Preprint posted online on Jun 20, 2025. [doi: 10.2139/ssrn.5268176]
65. Ignatowicz A, Tarrant C, Mannion R, El-Sawy D, Conroy S, Lasserson D. Organizational resilience in healthcare: a review and descriptive narrative synthesis of approaches to resilience measurement and assessment in empirical studies. BMC Health Serv Res 2023 Apr 19;23(1):376. [doi: 10.1186/s12913-023-09242-9] [Medline: 37076882]
66. Ratliff HC, Lee KA, Buchbinder M, Kelly LA, Yakusheva O, Costa DK. Organizational resilience in healthcare: a scoping review. J Healthc Manag 2025;70(3):165-188. [doi: 10.1097/JHM-D-24-00084] [Medline: 40358108]
67. Valiaiev D. Implementing DataOps: a scalable framework for modern data warehousing [PhD thesis]. : University of Arkansas at Little Rock; 2025 URL: https://search.proquest.com/openview/aa9acce88c488cbb8c6b3c8339062319/1 [accessed 2026-08-09]
68. Bharadwaj Mannapur S. Understanding data drift and concept drift in machine learning systems. Int J Sci Res Comput Sci Eng Inf Technol 2025;11(1):318-330. [doi: 10.32628/CSEIT25111239]
69. Betha R. Data observability and data quality automation: building self-healing data pipelines. J Adv Dev Res 2023;14(1):1-10. [doi: 10.5281/zenodo.15104120]
70. Kottam S, Annem S, Sun Y. Error-aware text-to-SQL generation for clinical trial eligibility criteria querying in EHR databases. In: 2025 IEEE International Conference on Big Data (BigData): IEEE; 2025. [doi: 10.1109/BigData66926.2025.11402102]
71. Alukal G, Manos A. Lean Kaizen: A Simplified Approach to Process Improvements: ASQ Quality Press; 2006.
72. UC Davis case study. HIMSS. URL: https://pages.himss.org/LP-HA-Case-Study-UC-Davis.html [accessed 2026-08-09]




XSL•FO
RenderX

73. Kumar R, Sporn K, Waisberg E, et al. Navigating healthcare AI governance: the comprehensive algorithmic oversight and stewardship framework for risk and equity. Health Care Anal 2025 Aug 13. [doi: 10.1007/s10728-025-00537-y] [Medline: 40802160]
74. Labkoff S, Oladimeji B, Kannry J, et al. Toward a responsible future: recommendations for AI-enabled clinical decision support. J Am Med Inform Assoc 2024 Nov 1;31(11):2730-2739. [doi: 10.1093/jamia/ocae209] [Medline: 39325508]
75. Loshin D. Master Data Management: Morgan Kaufmann; 2010.
76. Zimmermann S, Rentrop C, Felden C. A multiple case study on the nature and management of shadow information technology. J Inf Syst 2017 Mar;31(1):79-101. [doi: 10.2308/isys-51579]
77. Ipeirotis P, Zheng H. Natural language interfaces for databases: what do users think? arXiv. Preprint posted online on Nov 18, 2025 URL: https://arxiv.org/html/2511.14718v1 [accessed 2026-08-09]
78. Gartlehner G, Kugley S, Crotty K, et al. Artificial intelligence-assisted data extraction with a large language model: a study within reviews. Ann Intern Med 2025 Dec;178(12):1763-1771. [doi: 10.7326/ANNALS-25-00739] [Medline: 41183336]
79. Benzarti S, Berrabah C. Generative AI for intelligent data extraction: a case study in automated Excel-to-SQL with human oversight. In: Emerging Technologies for Developing Countries: Springer; 2026:151-164. [doi: 10.1007/978-3-032-16638-8_11]
80. Ning Z, Tian Y, Zhang Z, Zhang T, Li TJ. Insights into natural language database query errors: from attention misalignment to user handling strategies. ACM Trans Interact Intell Syst 2024 Dec;14(4):1-32. [doi: 10.1145/3650114]
81. Francisco Partners to acquire IBM's healthcare data and analytics assets. IBM. 2022. URL: https://newsroom.ibm.com/2022-01-21-Francisco-Partners-to-Acquire-IBMs-Healthcare-Data-and-Analytics-Assets [accessed 2026-08-09]
82. Strickland E. IBM Watson, heal thyself: how IBM overpromised and underdelivered on AI health care. IEEE Spectr 2019;56(4):24-31. [doi: 10.1109/MSPEC.2019.8678513]
83. Son H. Haven, the Amazon-Berkshire-JPMorgan venture to disrupt health care, is disbanding after 3 years. CNBC. URL: https://www.cnbc.com/2021/01/04/haven-the-amazon-berkshire-jpmorgan-venture-to-disrupt-healthcare-is-disbanding-after-3-years.html [accessed 2026-08-09]
84. Acchiardo JM, Gunderman RB. The failure of Haven Healthcare: lessons for radiology learners. Acad Radiol 2021 Jul;28(7):1029-1031. [doi: 10.1016/j.acra.2021.03.020] [Medline: 33896719]
85. Harrold ST. Health care analytics challenges: a 3-pillar framework connecting analytics maturity, workforce agility, and technical enablement. Zenodo. URL: https://doi.org/10.5281/zenodo.18264359 [accessed 2026-08-13]


## Abbreviations

**AMAM:** Analytics Maturity Assessment Model
**EMR:** electronic medical record
**HIMSS:** Healthcare Information and Management Systems Society
**HITL-KG:** human-in-the-loop knowledge governance
**LLM:** large language model
**NL2SQL:** natural language to SQL


*Edited by M Balcarras; submitted 30.Mar.2026; peer-reviewed by M Hamedani, X Zhang; revised version received 11.Jul.2026; accepted 13.Jul.2026; published 31.Aug.2026.*

*Please cite as:*
*Harrold ST*
*Health Care Analytics Challenges: A 3-Pillar Framework Connecting Analytics Maturity, Workforce Agility, and Technical Enablement*
*Interact J Med Res 2026;15:e96541*
*URL: https://www.i-jmr.org/2026/1/e96541*
*doi:10.2196/96541*


© Samuel Thomas Harrold. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 31.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

ViewPoint

# Evidence-Based Self-Management Strategies for Fibromyalgia: Foundations for Digital Therapeutic Applications

Thomas Lebeau Foustoukos[1,2], MSc; Isabelle Leclercq[2], MSc; Marc Blanchard[2], PhD; Thomas Hügle[2], MD, PhD

[1]College of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh, United Kingdom

[2]Department of Rheumatology, University Hospital of Lausanne, Lausanne, Vaud, Switzerland

**Corresponding Author:**
Thomas Hügle, MD, PhD
Department of Rheumatology
University Hospital of Lausanne
Av. Pierre-Decker 4
Lausanne, Vaud, 1011
Switzerland
Phone: 41 21 314 14
Email: thomas.hugle@chuv.ch

## Abstract

Fibromyalgia is a prevalent musculoskeletal pain condition that causes major personal, social, and societal burden. Pharmacological therapies often provide only limited benefit, making multimodal approaches and self-management the cornerstones of care. Such strategies, spanning lifestyle modification, physical activity, psychoeducation, and cognitive-behavioral approaches, target the biopsychosocial complexity of fibromyalgia and promote sustainable coping. In parallel, digital health technologies are transforming how these interventions can be delivered and coordinated in the form of digital therapeutics. This viewpoint draws on a multiphase investigation to appraise the current and future landscape of fibromyalgia self-management in the digital era. Its objective is to present an evidence-based framework and recommendations to guide the development of a mobile health self-management program for patients with fibromyalgia. In phase 1, we conducted a review of international guidelines and randomized controlled trial–based systematic reviews addressing nondigital self-management interventions for fibromyalgia and related nociplastic pain conditions. In phase 2, we analyzed the content and certification status of currently available mobile and virtual health applications for fibromyalgia. In phase 3, we convened a multidisciplinary focus group of rheumatologists, patients, and digital health developers to identify priorities for translating evidence-based self-management content into mobile health formats. Collectively, we suggest that effective digital self-management for fibromyalgia should evolve beyond single-domain interventions toward validated, personalized, and interactive multimodal platforms. Virtual care may increasingly function at the point of care, linking monitoring, education, and behavioral support in one continuum.

***(Interact J Med Res 2026;15:e67523)*** doi:10.2196/67523

**KEYWORDS**

chronic pain; cognitive behavior therapy; digital therapeutics; eHealth; fibromyalgia; mind-body therapies; mobile application; physical activity; self-management

## Introduction

Fibromyalgia is defined as a syndrome characterized by moderate to severe symptoms, including widespread pain, fatigue, sleep disturbance, cognitive complaints, and an increase in somatic complaints [1]. Despite ongoing efforts to clarify the definition of fibromyalgia, accurately assessing its diagnosis, prevalence, management, and its personal and societal impacts remains challenging [2]. Fibromyalgia can be accompanied by psychiatric, rheumatological, gastrointestinal, or other comorbidities, which further complicate the management and contribute to its overall personal and societal burden [3].

Proposing a clear treatment strategy for fibromyalgia remains challenging. The focus is on nonpharmacological therapies, particularly physical interventions, psychological treatments, and mind-body approaches that aim to promote self-management. Patient education plays a crucial role in enhancing their understanding of the disease and fostering their active participation in its management. Pharmacological treatments have shown limited benefits [4-8]. Multimodal treatment programs are recommended for treatment-resistant



XSL•FO
RenderX

fibromyalgia patients, as they combine different treatment modalities and offer group effects. However, such programs require significant resources that are not necessarily considered cost-effective.

Self-management programs play a crucial role in supporting individuals with chronic conditions [9] and are recommended by management fibromyalgia guidelines [4-7]. The fundamental concept of self-management emphasizes the active involvement of patients [9,10]. Self-management considers the multidimensional nature of the illness and combines multiple interventions such as psychological interventions, exercise, and nutritional aspects. It offers a relevant approach to address the challenges in fibromyalgia management and shows its efficiency to improve fibromyalgia patients' well-being [11,12] and a reduced tendency of patients to seek repeated consultations with health care professionals. A plethora of these programs have shown an improvement in pain intensity, functionality, cognitive behavior, emotions, and quality of life (QoL). However, the quality of these studies is heterogeneous [13-17]. When programs are detailed, their references are often unclear, relying on peer opinions or interviews [18-24], or they may offer only one component, such as psychoeducation or exercises [25-28].

The digital transformation is continuously shaping the health care landscape. It encompasses various digital innovations such as online programs, apps, gaming, virtual reality, or, more recently, generative artificial intelligence (AI) and chatbots [29]. In a clinical context, mobile health (mHealth) apps have demonstrated several key functionalities, including diagnostics and clinical decision support, behavior change interventions, and the delivery of disease-related education and communication support. The latter 3 functionalities are particularly relevant for delivering self-management interventions, empowering patients to monitor and manage their own health conditions, and providing an additional avenue for health care professionals to support their patients' well-being [30]. In 2019, a law was passed in Germany that digital health applications are reimbursable by health insurance companies under certain conditions. Currently, 8 digital health applications for musculoskeletal indications are registered; 2 of those are for chronic pain syndromes, including fibromyalgia [31]. There was a second acceleration of eHealth during the COVID-19 pandemic, notably concerning telemonitoring and video consultations. In the meantime, eHealth interventions have been shown to enhance patients' QoL and treatment experiences [32,33].

Online programs and mobile apps have demonstrated promising benefits for individuals with chronic pain or musculoskeletal conditions [33,34]. The use of mHealth technologies offers an accessible and available 24/7 solution without geographical constraints, empowering patients to become more self-reliant in managing their chronic conditions [35,36]. Notwithstanding, there are several caveats and limitations in existing solutions. A substantial proportion of currently available online fibromyalgia programs are not evidence-based or lack proven, evidence-based content. They are usually not personalized and cannot be updated, as this would alter the certification status of the device. On the other hand, recent digital tools such as large language models (LLMs) and chatbots offer novel approaches that could address some of these limitations.

## Aim

The objective of this work is to present an evidence-based framework and recommendations to guide the development of an mHealth self-management program for patients with fibromyalgia. A comprehensive literature review was conducted to evaluate the effectiveness of self-management interventions on symptoms in patients with fibromyalgia (phase 1). Where applicable, relevant practical characteristics of the interventions were also reported. In phase 2, existing therapeutic fibromyalgia apps and online programs were reviewed, and in phase 3, general aspects and user preferences were discussed in focus groups with patients. Finally, we aimed to identify strategies to optimize online self-management interventions and explore their potential for further improvement.

## Ethical Considerations

This study is based on previously conducted studies and does not contain any new studies with human participants or animals performed by any of the authors.

## Phase 1: Review of Existing Evidence

### Overview

MEDLINE and Cochrane databases, encompassing English and French articles from the inception of the databases up until 2023, were searched to identify guidelines and systematic reviews of randomized controlled trials for fibromyalgia self-management. Data from 30 systematic reviews that demonstrated a high or moderate level of confidence based on the AMSTAR 2 evaluation and 8 guidelines with an AGREE II–GRS score higher than 18 were considered to categorize and summarize the evidence on self-management interventions (Figure S1 in Multimedia Appendix 1).

### Categorization of Self-Management Interventions

Self-management interventions were categorized into 3 groups: physical exercise, psychoeducative interventions, and mind-body interventions. Additionally, a separate category called multimodal interventions was identified, which encompassed interventions combining 2 or more of the previously mentioned self-management approaches. The European, American, Canadian, and Italian guidelines covered physical exercise, psychoeducative interventions, and mind-body self-management interventions as part of the overall management of patients with fibromyalgia, alongside other treatment modalities [4,6,7,37]. The German guideline divided its recommendations into 4 separate studies: general recommendations [7], recommendations on physical interventions [38], psychological interventions [39], and other interventions such as mind-body therapies [40]. In total, 8 studies focused on physical activities [41-48]; 14 studies on psychoeducative interventions [25,49-60]; 4 studies on mind-body therapies [61-64]; 1 study on multicomponent interventions [11]; 1 study that investigated both psychoeducative and mind-body interventions [65]; and 2



XSL•FO
RenderX

studies that examined physical activities, psychoeducative interventions, mind-body therapies, and multicomponent interventions [66,67]. The studies used various types of control groups, which encompassed the following options: treatment as usual, waiting list or no treatment, alternative pharmacological or nonpharmacological interventions, active or attention control, and placebo or sham interventions. Designing a control group for complex self-management interventions was considered challenging due to the inherent difficulty of blinding participants, which can introduce measurement and interpretation bias [68].

### Evidence for Specific Self-Management Interventions

A wide range of variables were assessed both before and after the interventions: pain (both qualitative and quantitative measures), QoL, psychological functions such as depression, anxiety, self-efficacy, acceptance, catastrophizing, and fear-avoidance, physical functions such as strength and disability, sleep quality, fatigue, health care use indicators such as visits, sick leaves, return to work, and care-seeking behavior, as well as trial withdrawal rates and adverse effects. The Fibromyalgia Impact Questionnaire (FIQ) was used as a multidimensional measurement tool.

### Physical Interventions

All of the guidelines strongly endorse the inclusion of aerobic exercises and resistance (or strength) exercises in self-management interventions for fibromyalgia [4,6,7,37,38]. On the other hand, flexibility and stretching exercises are either not recommended [6,7] or recommended with weak or insufficient evidence of benefit [37,38]. Exercise interventions were found to be superior to control groups in improving pain, FIQ scores, sleep quality, fatigue, and depression [67]. The characteristics of the exercises were extracted based on the prescription guidelines provided by the American College of Sports Medicine, which include frequency, intensity, time, and type of exercises [69]. Additionally, the duration of the exercises, delivery modalities, whether they were tailored or standardized, and safety considerations were also considered. Aerobic exercises had positive effects on pain, QoL, and depression [42,45,46]. Aerobic exercises demonstrated a high effect on reducing the FIQ [41]. Aerobic exercises probably improve stiffness and slightly improve physical function and cardiovascular function [42]. According to Couto et al [46], resistance exercises yielded positive effects on pain and QoL but did not show a significant effect on depression. On the other hand, Albuquerque et al [41] found that resistance exercise demonstrated a moderate effect in reducing the FIQ scores. Resistance exercise has the potential to improve fatigue and sleep [49]. Although flexibility exercise did not show superiority over control groups (other interventions or no intervention), the quality of evidence is insufficient to draw definitive conclusions regarding the effectiveness of flexibility exercise on fibromyalgia symptoms [41,45,46,48]. However, there is evidence suggesting that physical exercise may help decrease muscle stiffness [48]. Interventions that combine aerobic exercises, resistance exercises, and flexibility exercises have shown greater effectiveness in improving FIQ scores compared to single-type exercise interventions [41]. However, Bidonde et al [43] state that no definitive conclusion can be drawn regarding the optimal proportion, synergy, or specific characteristics of each exercise type [43]. Exergames have demonstrated promising benefits in improving pain, disability, and physical function in both the short and long term [66]. There is a lack of available information regarding other exercises, such as Pilates and motor control exercises. Exercise (aerobic, resistance, and flexibility) characteristics are shown in Table S1 in Multimedia Appendix 1.

### Psychoeducative Interventions

Consensus guidelines recommend the inclusion of psychological and educational interventions, such as cognitive behavioral therapy (CBT) and health education [4-7,37]. These interventions can encourage self-management, enhance self-efficacy, and reduce maladaptive thoughts and behaviors [6]. It is particularly important to consider these interventions when patients have comorbid mental disorders [39]. For hypnosis and guided imagery, recommendations are conflicting [4,7,39]. Although other psychological interventions, such as relaxation, therapeutic writing, Roger therapy, family therapy, psychodynamic therapy, and psychoanalytic therapy, are available, they should not be proposed as stand-alone treatments due to the lack of evidence or limited recommendations [39]. CBT incorporates education on the physiopathology of pain, self-management skills, cognitive reappraisal, pacing activities, and problem-solving techniques to help patients modify their behaviors, thoughts, and emotions, ultimately reducing pain, improving functioning, and enhancing mood. Additional components, such as sleep hygiene, may be included. Homework assignments are given to encourage the practice of skills in everyday life [60]. Acceptance and commitment therapy (ACT) and mindfulness-based therapy are extensions of CBT. ACT comprises 2 main core components: mindfulness and acceptance, as well as commitment and behavior change. It involves several processes, including acceptance, cognitive defusion, present-moment awareness, self-as-context, values clarification, and committed action [70]. ACT interventions often incorporate CBT processes, making it difficult to distinguish between CBT and ACT [51]. Evidence suggests that CBT can reduce pain, disability, and negative mood in patients with chronic pain or fibromyalgia, both immediately after treatment and at long-term follow-up (6 and 12 months), although the effect size is small or very small [68,79. Enomoto et al [52] recommend offering CBT for insomnia, or at least CBT for insomnia and pain, to patients with chronic pain who also experience insomnia as a comorbidity. CBT therapy for insomnia has been effective in improving sleep, pain, disability, and depression. ACT has also demonstrated benefits for patients with chronic pain, improving pain acceptance, QoL, pain intensity, functioning, and mood [57,59]. For fibromyalgia, ACT has shown significant improvements in patient functioning in both the short and long term, based on a meta-analysis with moderate-quality evidence [51]. ACT can be considered to enhance patients' psychological flexibility and, subsequently, their functioning in pursuing valued activities [51,59]. Education serves as both an intervention and a CBT tool [58]. Joypaul et al [55] defined education as providing instructions to inform participants, making it applicable to various interventions as an instructional



XSL•FO
RenderX

tool [55]. Therapeutic pain neuroscience education, also referred to as pain neuroscience education by some authors, aims to enhance patients' knowledge and understanding of pain neurophysiology to improve their cognitive and behavioral skills related to pain [56,71]. Therapeutic pain neuroscience education has demonstrated benefits for patients with chronic pain in reducing fear of movement, pain intensity, pain disability, and pain catastrophizing [58]. Therapeutic pain neuroscience education, in the context of chronic musculoskeletal pain, improved pain intensity, pain knowledge, disability, maladaptive thoughts and behaviors, physical function, and health care use, even up to 1 year post treatment [56]. Other educational content has been studied, including lifestyle components focusing on areas such as nutrition, sexuality, social coping strategies, and the regulation and adjustment of everyday life [11]. Education should be an integral part of a multidisciplinary approach to chronic pain, alongside graded physical activities, graded exposure, and pacing, for example, and not offered as a stand-alone treatment [54-56,58,67]. CBT education characteristics were gathered in Table S1 in Multimedia Appendix 1.

### Mind-Body Interventions

The classification of mind-body interventions was based on the Medical Subject Headings terms of the National Library of Medicine. They encompass various modalities such as meditative movement therapies (MMTs), respiration exercises, hypnosis and autogenic training, meditation, and relaxation [72]. They focus on exploring the interconnectedness between the brain, body, mind, and behavior, and how emotional, mental, social, spiritual, experiential, and behavioral factors can directly impact health [73]. Guidelines recommend the use of MMTs with confidence [4,7,40]. However, hypnosis, guided imagery, relaxation, or meditation are not recommended, and it is advised not to propose them as stand-alone interventions [7,40]. The Canadian guideline suggests not discontinuing these interventions but informing the patient about their lack of evidence and potential side effects [6]. Therapies such as Qi Gong, Tai Chi, and Yoga, as examples of MMTs, were included [63]. MMTs have shown improvements in outcomes such as physical functioning, pain, and mood for patients with fibromyalgia [65]. Specifically, Tai Chi has been found to significantly reduce FIQ scores, pain intensity, sleep disturbance, fatigue, and depression, while increasing QoL for individuals with fibromyalgia [61]. Tai Chi and Yoga are weakly recommended, whereas the evidence did not allow for a recommendation regarding Qi Gong for chronic pain management [63]. Lee et al [64] did not provide a recommendation for meditation as a treatment for chronic pain due to the limited evidence available. While no specific recommendation was provided in the systematic review, Lee et al [64] suggested that mindfulness-based interventions could potentially offer benefits for chronic pain management.

### Multicomponent Interventions

Multicomponent interventions were defined as combinations of at least 2 components, including psychological, physical activity, medical education, and mind-body therapies, based on the self-management program developed by Miles et al [74] and further refined by Geraghty et al [11]. All guidelines recommend not implementing interventions in isolation but rather combining them with other approaches. Multicomponent intervention should include physical and psychoeducational interventions at a minimum [4-7,75]. Multicomponent interventions have demonstrated effectiveness in improving various aspects such as FIQ, pain, sleep, and depression, with greater effects compared to exercise alone, education alone, or psychological interventions alone [61,62]. The positive effects of multicomponent interventions typically last for an average of 14 weeks, and it is recommended to conduct follow-up assessments every 3 months to review and reinforce the treatment strategy [62]. Considering the variations across studies, multicomponent interventions have shown improvements in physical function, pain, FIQ, fatigue, mood, and QoL in both the short and long term [11].

## Phase 2: Review of Self-Management Apps

In this part of the work, we investigated which type of self-management is integrated in existing apps for fibromyalgia and chronic pain syndromes (Table 1). Online fibromyalgia applications were identified via PubMed, Google, and the German Digitale Gesundheitsanwendung (DiGA) registry [76], as well as with ChatGPT (OpenAI). Available mobile apps targeting fibromyalgia self-management reveal a strong emphasis on psychoeducational content, particularly CBT, ACT, and pain neuroscience education. Out of the 12 reviewed apps, 10 integrate psychoeducational modules, often presented through structured lessons, interactive exercises, journaling tools, or virtual coaching. Physical activity interventions, such as aerobic or resistance exercise guidance, were integrated in 6 apps, typically via instructional videos, activity logging, or synchronization with wearable devices. Mind-body techniques, including guided meditation, breathing exercises, and mindfulness-based practices, were found in about 5 apps. This distribution highlights a predominant focus on psychological and cognitive strategies in current digital interventions, with fewer apps addressing physical or somatic components in depth.



XSL•FO
RenderX

**Table 1.** Existing online self-management programs for fibromyalgia.

| App name | Therapeutic mechanisms | Regulatory status | Notes |
|---|---|---|---|
| Stanza | ACT[a], CBT[b], mindfulness, journaling, reminders | FDA[c]-cleared in the United States | 12-week structured digital therapy, clinical trial support |
| HelloBetter Chronic Pain | CBT, ACT, mindfulness | DiGA[d]-approved in Germany | Structured 12-week course, reimbursed by German health insurance |
| Selfapy Chronic Pain | CBT, mindfulness, education | DiGA approval in Germany withdrawn | Evidence-based course, reimbursed via DiGA |
| Manage My Pain | Pain tracking, journaling, self-monitoring | Available on app stores | Focuses on tracking pain trends for self-awareness and provider communication |
| Quell Fibromyalgia | Neuromodulation via wearable, symptom tracking | FDA-authorized medical device | App supports wearable neuromodulation therapy for fibromyalgia |
| FibroMapp | CBT-based tools, symptom tracking, medication management | Available on app stores | Combines tracking and CBT tools; user-driven format |
| MoreGoodDays | CBT, mindfulness, education | Available on app stores | Structured CBT modules, lifestyle education, mindfulness tools |
| FibroMinder | Reminders, symptom tracking, task management | Available on app stores | Simple utility app for scheduling and symptom tracking |
| PainScale | Pain tracking, education, CBT elements, community support | Available on app stores | Community-focused platform with integrated education and logging |
| FibroTrack | Symptom tracking, lifestyle management | Available on app stores | Focuses on tracking and lifestyle planning features |
| Curable | Pain neuroscience education, CBT, guided meditation, writing exercises | Available on app stores | Comprehensive app targeting pain beliefs and coping strategies |
| Fibrowalk | CBT, mindfulness, therapeutic exercise, pain neuroscience education | Not a certified app; delivered via YouTube and email | Multicomponent program with weekly video sessions |

[a]ACT: acceptance and commitment therapy.

[b]CBT: cognitive behavioral therapy.

[c]FDA: US Food and Drug Administration.

[d]DiGA: Digitale Gesundheitsanwendung.

Two applications, HelloBetter Chronic Pain and Stanza, are certified digital therapeutics. HelloBetter is DiGA-approved in Germany, meaning they are reimbursable by statutory health insurance and have met criteria for clinical evidence and data security. Stanza, developed in the United States, is US Food and Drug Administration–cleared as a prescription digital therapeutic, offering a 12-week ACT-based program with demonstrated clinical efficacy in reducing fibromyalgia symptoms [77]. The remaining apps, including Curable, MoreGoodDays, Manage My Pain, and FibroMapp, are publicly available but lack medical certification and are typically framed as wellness tools or digital companions rather than regulated treatments. Overall, while several apps offer well-designed psychoeducational modules and basic symptom tracking, fewer integrate evidence-based physical activity interventions or comprehensive mind-body components. Among the 12 fibromyalgia self-management apps analyzed, only 5 (42%) can be considered multicomponent, meaning they integrate at least 2 of the 3 core evidence-based intervention domains: psychoeducational therapies, physical activity, and mind-body techniques. Notably, the certified apps HelloBetter Chronic Pain, Selfapy Chronic Pain, and Stanza fall into this category, offering combinations of CBT or ACT, mindfulness, and, in some cases, activity planning. Noncertified apps such as MoreGoodDays and Curable also include psychoeducation and mind-body interventions, but vary in the depth and clinical rigor of their content. While most of the reviewed apps integrate evidence-based components aligned with current clinical guidelines, such as CBT, ACT, and mindfulness-based techniques, several also include strategies with limited or inconclusive evidence in fibromyalgia. For example, some noncertified apps incorporate generic meditation, unstructured journaling, expressive writing, or broad lifestyle advice without therapeutic framing or individual tailoring.

## Phase 3: Patient Preferences and Focus Groups

As previously reported, we investigated patient preferences for mobile app design through 2 online surveys (53 and 33 patients, respectively) and 3 focus groups comprising a rheumatologist, 1-6 fibromyalgia patients, with or without concomitant post-COVID-19 syndrome, as well as an app designer [78]. All patients in focus groups tested a self-developed app designed to assess fibromyalgia symptoms and to deliver an online therapeutic program containing CBT, mind-body techniques, and physical exercise instructions. This program was initially developed during the pandemic for fibromyalgia associated with postviral syndrome and was subsequently adapted as an



XSL•FO
RenderX

experimental chronic pain companion (Pain Organiser and Companion System). The design and content of this app are described in detail elsewhere [78,79]. In the final focus group (n=1), the online program was connected to an LLM-based chatbot and tested by the patient (see Future Directions section).

Participants emphasized that simplicity, clarity, and accessibility are crucial for adherence, given common cognitive fatigue ("fibrofog") and limited digital literacy. Users preferred short, well-structured modules, clear navigation, and scientifically validated content in plain, nontechnical language. Autonomy and empowerment emerged as central themes. Patients valued tools that visualize progress, such as symptom tracking and diaries, helping them recognize links between behavior and symptoms. Personalization was highly desired; users wanted adaptable exercises, favorite lists, and an interactive virtual coach or chatbot that responds to individual needs and previous activity. An empathic, motivating tone was considered essential. Participants appreciated friendly communication and multimedia content, particularly physiotherapy videos and mindfulness modules. Emotional design elements, positive wording, supportive colors, and encouraging feedback were viewed as vital for engagement and adherence. Reliability and trustworthiness were further priorities: users expected apps to be evidence-based, transparent, and secure, with clear data protection and professional validation. Barriers such as poor navigation, cognitive overload, or lack of emotional connection can significantly reduce adherence. The focus group highlighted that fibromyalgia mHealth apps offer the potential to serve as multicomponent platforms, covering the 3 key therapeutic domains discussed above: psychoeducation, mind-body techniques, and physical exercise guidance. Of note, instructions for physical exercises were appreciated by participants in the focus groups, ideally with more individualization (eg, for individuals with obesity).

## Key Findings and Lessons Learned

There is proven evidence for psychoeducational interventions, mind-body therapies, and physical activities in fibromyalgia. All 3 are included in most of the current fibromyalgia mHealth applications on the market, several of those with proven efficacy in randomized controlled trials, although the underlying therapeutic content is not openly accessible [75]. CBT remained strongly recommended to assist patients with fibromyalgia in managing their pain, comorbidities, and lifestyle [49] and is the main approach used in existing fibromyalgia apps. Apart from CBT interventions, psychoeducation focuses on pain coping skills, mindfulness, lifestyle modification, and pain neuroscience education. The recommendations regarding mind-body therapies such as MMT, meditation, relaxation, or hypnosis are conflicting but safe and therefore might be integrated in apps as an adjunct. Certain interventions, such as stress reduction, breathing, and relaxation exercises, are notably suitable for integrating into apps as animation can be positive, such as a balloon for breathing exercises.

Half of the existing fibromyalgia apps integrate physical exercises into their program, but mostly only to a lower degree compared to CBT. The content of existing apps contrasts with technically more sophisticated physiotherapy or online rehabilitation apps; for example, integrating computer vision to instruct and monitor a wide range of exercises. In general, it can be postulated that true multicomponent apps integrating all 3 validated pillars of fibromyalgia self-management, psychoeducation, physical activity, and mind-body techniques, guided by clinical standards and codeveloped with patient input, would improve the effect.

From a technical side, both web and native apps are viable platforms for integrating self-management tools and tracking disease activity in fibromyalgia. Native apps are particularly effective at incorporating wearable data directly from mobile devices, enabling continuous monitoring and data capture. Symptom tracking is typically based on patient-reported outcomes or symptom scores, including measures of pain, QoL, depression, anxiety, self-efficacy, acceptance, catastrophizing, fear-avoidance, physical function (eg, strength and disability), sleep quality, fatigue, and health care use indicators such as clinic visits, sick leave, and return-to-work patterns.

These apps often incorporate validated instruments such as the FIQ, the Symptom Severity Score, or the Widespread Pain Index to assess disease activity. Therapeutic content is typically delivered through text, animated videos, or voice formats, providing education, instructions, and exercises. In some cases, apps are integrated with connected devices, such as wearables that deliver millimeter wave stimulation [80]. Animation can be used to enhance certain exercises, such as balloons being used to synchronize breathing frequency and intensity.

Stand-alone apps, particularly those certified as digital therapeutics, often include a structured program guide to track patient progress. This is important for potential reimbursement by health insurers. Increasingly, mHealth apps also integrate chatbots that guide users through the program, enhance flexibility, and potentially improve adherence, one of the main challenges of these interventions. Successful therapeutic apps for this population require a balance between scientific rigor, user-friendly design, and emotionally intelligent interaction. A clinical and evidence-based framework for fibromyalgia interventions in apps can be found in Table 2. Textbox 1 presents practical recommendations for the corresponding app design and implementation roadmap.



XSL•FO
RenderX

**Table 2.** Evidence-based framework for fibromyalgia online interventions.

| Intervention domain | Evidence level and rationale | Implementation in mobile health app | Digital enablement features | Evaluation metrics |
|---|---|---|---|---|
| Psychoeducational therapies (CBT[a], ACT[b], and PNE[c]) | Strong evidence from RCTs[d] and guidelines for reducing pain, distress, and maladaptive thoughts | Modular courses; microlearning sessions; quizzes; journaling; goal setting; and self-efficacy tracking | Chatbots or avatars for guidance; adaptive progression based on symptom and mood data; empathy-based conversational tone; and voice message transcription | Pain, FIQ[e], self-efficacy, catastrophizing, engagement rate, and module completion |
| Physical activity and graded exercise | Strong evidence for aerobic and resistance training improving function and mood | Video demonstrations; personalized exercise plans; daily movement reminders; wearable integration (steps, HR[f], and HRV[g]) | AI[h]-driven motion feedback (computer vision); adaptive load progression; safety alerts; and gamification | Physical function (FIQ physical domain), fatigue, adherence (logged sessions), and HR and HRV trends |
| Mind-body techniques (mindfulness, Tai Chi, breathing, and relaxation) | Moderate evidence for improving sleep, mood, and QoL[i] | Audio-guided meditations, breathing animations, mindfulness timers, and relaxation music | Adaptive session lengths; stress biofeedback using HRV; sleep tracking linkage; and integration with smartwatches | Sleep quality, stress index, anxiety, HRV, and app usage continuity |
| Symptom tracking and patient-reported outcomes | Essential for personalization and clinical insight for HCP[j] | In-app FIQ, pain diaries, and fatigue and sleep trackers | Voice or video symptom entry; AI-based summary visualization; and adaptive dashboard | Data completeness, trend accuracy, and correlation with clinical outcomes |
| Personalization and adaptive design | Increasingly essential for engagement and relevance | Custom goal setting; phenotype-based pathways (eg, obesity-, menopause-, and PTSD[k]-related FM[l]) | Machine learning–based tailoring; predictive suggestions for pacing and exercise. Adapted avatars, eg, older people with obesity | User satisfaction, engagement over time, and adaptive accuracy |
| Behavior change and motivation | Crucial for long-term adherence | SMART[m] goal planning, feedback loops, and progress visualization | Gamification, motivational messaging, positive reinforcement, and social comparison (optional) | Retention rate, adherence index, and self-efficacy gain |
| Communication and support | Improves adherence and patient safety | Chatbot or professional chat; asynchronous therapist feedback | Hybrid care integration (eg, AI triage + human follow-up); crisis escalation paths | Message frequency, satisfaction, and safety events |
| Data integration and clinical workflow | Enables clinical supervision and research | Clinician dashboard, FHIR[n]-based interoperability, and data export to EHR[o] | Sidecar EMR integration; secure teleconsultation channel | Clinical uptake, data completeness, and clinician feedback |
| Accessibility, UX[p], and emotional design | Critical for usability and adherence in cognitive fatigue | Clean interface, large icons, voice navigation, and light and dark modes | Emotionally supportive design (colors, feedback tone), simplified onboarding, and language localization | SUS[q] score, accessibility compliance, and dropout rate |
| Privacy, certification, and ethics | Required for trust and scalability | CE[r], FDA[s], or DiGA[t] conformity, and transparent data policies | Privacy-by-design architecture; on-device data processing | Certification status, GDPR[u] or HIPAA[v] compliance, and user trust rating |

[a]CBT: cognitive behavioral therapy.

[b]ACT: acceptance and commitment therapy.

[c]PNE: pain neuroscience education.

[d]RCT: randomized controlled trial.

[e]FIQ: Fibromyalgia Impact Questionnaire.

[f]HR: heart rate.

[g]HRV: heart rate variability.

[h]AI: artificial intelligence.

[i]QoL: quality of life.

[j]HCP: health care professional.

[k]PTSD: posttraumatic stress disorder.

[l]FM: fibromyalgia.

[m]SMART: specific, measurable, achievable, relevant, time-bound.

[n]FHIR: Fast Healthcare Interoperability Resources.



XSL•FO
RenderX

[o]EHR: electronic health record.

[p]UX: user experience.

[q]SUS: System Usability Scale.

[r]CE: Conformité Européenne.

[s]FDA: US Food and Drug Administration.

[t]DiGA: Digitale Gesundheitsanwendung.

[u]GDPR: General Data Protection Regulation.

[v]HIPAA: Health Insurance Portability and Accountability Act.

**Textbox 1.** Recommendations for the fibromyalgia mobile health self-management program development.

**Recommendation**

- Involve key stakeholders (fibromyalgia patients, multidisciplinary professionals, and digital experts) through surveys, interviews, and co-design workshops.
- Conduct a structured needs assessment, addressing core fibromyalgia symptoms and patient priorities (eg, fatigue, pain, and cognitive dysfunction).
- Provide a transparent overview of the content for users, health care providers, and regulators, eg, using knowledge graphs or content maps.
- Tailor content to user characteristics such as socioeconomic background, gender, age, culture, and health literacy.
- Base all interventions on clinical evidence, including physical activity, cognitive behavioral therapy, psychoeducation, and mind-body strategies.
- Include personalized modules for subtypes and comorbidity (eg, perimenopausal symptoms, obesity, anxiety, posttraumatic stress disorder–associated hypermobility, migraine, etc).
- Adhere to established standards and guidelines, such as Xcertia, National Institute for Health and Care Excellence, Haute Autorité de Santé, and World Health Organization, covering content, privacy, usability, and operability.
- Design the app as a complementary tool, not a substitute for face-to-face care.
- Use interoperable data formats, such as Fast Healthcare Interoperability Resources, to enable secure integration with health care systems and devices.
- Connect to wearable data, allowing tracking of physical activity, sleep, or heart rate variability to enrich outcome measures.
- Include a clinician-facing dashboard to enable remote monitoring, triage, and decision support.
- Ensure robust data security and privacy compliance, including user consent, encryption, and data governance.
- Provide a structured and simple onboarding process, guiding users through initial setup and goals.
- Ensure user-friendliness with intuitive design, a responsive user interface, customizable settings, reminders, and accessible language.
- Include motivational features, such as gamification, progress tracking, and positive reinforcement.
- Incorporate core functional modules, including symptom tracking, educational content, communication, and self-assessment tools.
- Implement adaptive, guided interventions rather than static or generic content to boost engagement and outcomes.
- Define and track recognized fibromyalgia end points to support evidence generation and reimbursement (eg, Fibromyalgia Impact Questionnaire, pain scales).
- Follow an iterative development cycle, including continuous evaluation of effectiveness, usability, and safety.
- Plan for scalability, dissemination, and regulatory approval, including Conformité Européenne marking, Digitale Gesundheitsanwendungen eligibility (Germany), or US Food and Drug Administration listing.

## Areas of Uncertainty

The literature search was limited to PubMed and Cochrane databases, potentially introducing selection bias, and the cutoff at 2023 excluded recent studies or emerging interventions. Another limitation of this viewpoint concerns the transferability of conventional evidence-based therapies to digital formats. The effectiveness of individual therapeutic elements within apps remains largely untested, and user experience likely plays a decisive role. Adherence is particularly critical, as unguided online interventions often show low engagement. Psychiatric comorbidities, such as depression, trauma, and cognitive fatigue, further affect adherence and suggest that some users may benefit from guided or coach-assisted programs rather than fully automated ones. The human and group-based components of multimodal therapy are difficult to replicate digitally, raising questions about the optimal degree of human involvement in digital therapeutics. Device-based or biofeedback interventions were not included, as they require additional hardware not universally accessible; however, future versions may integrate smartwatch-based or wearable data.

Finally, personalization remains a key challenge. As fibromyalgia is heterogeneous, tailored approaches—whether based on symptom profiles or machine learning–derived phenotypes—are needed. Future studies should also account for contextual factors such as age, culture, digital literacy, and



XSL•FO
RenderX

socioeconomic status, as well as identify which behavior change techniques drive engagement and outcomes. Guided and adaptive interventions appear more effective than static content, highlighting the importance of iterative, patient-centered design in future mHealth development.

## Future Directions

As seen in other digital interventions, such as for depression, the inclusion of a channel for direct interaction with a health care professional or, potentially, a bot can increase effectiveness. Voice message transcription features can facilitate semiautomated, asynchronous communication. In this context, agentic AI trained on such interactions may provide supplementary or alternative support. Dashboards play a critical role in integrating fibromyalgia mHealth applications into the clinical workflow, enabling health care professionals to monitor their patients and app activity. Ideally, these dashboards should interface with electronic medical records through sidecar applications [81]. Personalization of content according to fibromyalgia subtypes should be encouraged. However, this increased individualization may make the creation of robust evidence or standardized certification more challenging. As a future perspective, new content formats could be developed and evaluated, such as therapeutic stories or short interventions. For example, therapeutic stories could resemble YouTube Shorts, with AI algorithms applied to tailor and deliver these formats accordingly. Other future digital directions for fibromyalgia apps will move beyond psychoeducation toward multimodal, personalized therapy ecosystems. Integration of exercise modules, CBT, and mindfulness into cohesive platforms is expected, with adaptive interfaces that adjust content to each user's symptom patterns, motivation, and cognitive load. LLMs and AI-powered chatbots will evolve from static text-based assistants to emotionally intelligent digital companions that guide users through daily self-management, provide empathetic feedback, and interface with sensor data from wearables for sleep, mobility, or stress tracking. These systems could enhance adherence, detect flares early, and personalize pacing and exercise recommendations. Meanwhile, interoperability, data privacy, and regulatory validation will remain crucial to ensure clinical reliability and ethical use. Ultimately, the future of fibromyalgia apps lies in hybrid digital care, combining automated AI-driven support with human coaching and clinical oversight. By blending evidence-based therapy, personalization, and human empathy, next-generation digital therapeutics may finally bridge the gap between self-management and sustained symptom improvement.

## Conclusion

Evidence-based foundations are needed to inform the development of effective digital therapeutic applications for fibromyalgia. To build impactful digital therapeutics, it is essential to integrate validated self-management strategies, engage key stakeholders, and consider social and cultural determinants. An iterative development process, guided by ongoing assessment of usability, effectiveness, equity, and safety, will ensure that mHealth tools align with both clinical goals and patient needs.

### Acknowledgments

We thank Physio Homecare for providing physical exercise videos for the Pain Organiser and Companion System app. Generative artificial intelligence was used for editing the manuscript. Support for content creation is provided in Table 2.

### Data Availability

All data generated or analyzed during this study are included in this published article. Quantitative and qualitative data of focus groups can be found in [78].

### Authors' Contributions

All authors have contributed to the conception of this article and approved the final version to be published. TLF and TH wrote the manuscript. TLF performed the literature review and was involved in a focus group. MB and IL facilitated the focus groups. MB created the app and led focus groups. TH was involved in app development and led focus groups.

### Conflicts of Interest

MB and TH are the board members of Atreon.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 181 KB - ijmr_v15i1e67523_app1.docx ]

### References

1. Wolfe F, Clauw DJ, Fitzcharles M, Goldenberg DL, Häuser W, Katz RL, et al. 2016 revisions to the 2010/2011 fibromyalgia diagnostic criteria. Semin Arthritis Rheum 2016;46(3):319-329. [doi: 10.1016/j.semarthrit.2016.08.012] [Medline: 27916278]
2. Arnold LM, Bennett RM, Crofford LJ, Dean LE, Clauw DJ, Goldenberg DL, et al. AAPT diagnostic criteria for fibromyalgia. J Pain 2019;20(6):611-628 [FREE Full text] [doi: 10.1016/j.jpain.2018.10.008] [Medline: 30453109]



XSL•FO
RenderX

3. Lichtenstein A, Tiosano S, Amital H. The complexities of fibromyalgia and its comorbidities. Curr Opin Rheumatol 2018;30(1):94. [doi: 10.1097/bor.0000000000000464]
4. Ariani A, Bazzichi L, Sarzi-Puttini P, Salaffi F, Manara M, Prevete I, et al. The Italian society for rheumatology clinical practice guidelines for the diagnosis and management of fibromyalgia best practices based on current scientific evidence. Reumatismo 2021;73(2):89-105 [FREE Full text] [doi: 10.4081/reumatismo.2021.1362] [Medline: 34342210]
5. Eich W, Häuser W, Arnold B, Bernardy K, Brückle W, Eidmann U, Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. Fibromyalgia syndrome. General principles and coordination of clinical care and patient education. Schmerz 2012;26(3):268-275. [doi: 10.1007/s00482-012-1167-z] [Medline: 22760459]
6. Fitzcharles MA, Ste-Marie PA, Goldenberg DL, Pereira JX, Abbey S, Choinière M, et al. 2012 Canadian guidelines for the diagnosis and management of fibromyalgia syndrome: executive summary. Pain Res Manag 2013;18(3):119-126 [FREE Full text] [doi: 10.1155/2013/918216] [Medline: 23748251]
7. Macfarlane GJ, Kronisch C, Dean LE, Atzeni F, Häuser W, Fluß E, et al. EULAR revised recommendations for the management of fibromyalgia. Ann Rheum Dis 2017;76(2):318-328 [FREE Full text] [doi: 10.1136/annrheumdis-2016-209724] [Medline: 27377815]
8. Sarzi-Puttini P, Giorgi V, Marotto D, Atzeni F. Fibromyalgia: an update on clinical characteristics, aetiopathogenesis and treatment. Nat Rev Rheumatol 2020;16(11):645-660. [doi: 10.1038/s41584-020-00506-w] [Medline: 33024295]
9. Newman S, Steed L, Mulligan K. Self-management interventions for chronic illness. Lancet 2004;364(9444):1523-1537. [doi: 10.1016/S0140-6736(04)17277-2] [Medline: 15500899]
10. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
11. Geraghty AWA, Maund E, Newell D, Santer M, Everitt H, Price C, et al. Self-management for chronic widespread pain including fibromyalgia: a systematic review and meta-analysis. PLoS One 2021;16(7):e0254642 [FREE Full text] [doi: 10.1371/journal.pone.0254642] [Medline: 34270606]
12. Häuser W, Bernardy K, Arnold B, Offenbächer M, Schiltenwolf M. Efficacy of multicomponent treatment in fibromyalgia syndrome: a meta-analysis of randomized controlled clinical trials. Arthritis Rheum 2009;61(2):216-224. [doi: 10.1002/art.24276] [Medline: 19177530]
13. Cardosa M, Osman ZJ, Nicholas M, Tonkin L, Williams A, Abd Aziz K, et al. Self-management of chronic pain in Malaysian patients: effectiveness trial with 1-year follow-up. Transl Behav Med 2012;2(1):30-37 [FREE Full text] [doi: 10.1007/s13142-011-0095-2] [Medline: 22448204]
14. Cedraschi C, Desmeules J, Rapiti E, Baumgartner E, Cohen P, Finckh A, et al. Fibromyalgia: a randomised, controlled trial of a treatment programme based on self management. Ann Rheum Dis 2004;63(3):290-296 [FREE Full text] [doi: 10.1136/ard.2002.004945] [Medline: 14962965]
15. Glenn B, Burns JW. Pain self-management in the process and outcome of multidisciplinary treatment of chronic pain: evaluation of a stage of change model. J Behav Med 2003;26(5):417-433. [doi: 10.1023/a:1025720017595] [Medline: 14593851]
16. Rod K. Finding ways to lift barriers to care for chronic pain patients: outcomes of using internet-based self-management activities to reduce pain and improve quality of life. Pain Res Manag 2016;2016:8714785 [FREE Full text] [doi: 10.1155/2016/8714785] [Medline: 27445632]
17. Turner BJ, Rodriguez N, Bobadilla R, Hernandez AE, Yin Z. Chronic pain self-management program for low-income patients: themes from a qualitative inquiry. Pain Med 2020;21(2):e1-e8. [doi: 10.1093/pm/pny192] [Medline: 30312459]
18. LeFort SM, Gray-Donald K, Rowat KM, Jeans ME. Randomized controlled trial of a community-based psychoeducation program for the self-management of chronic pain. Pain 1998;74(2-3):297-306. [doi: 10.1016/s0304-3959(97)00190-5] [Medline: 9520244]
19. Lorig KR, Sobel DS, Stewart AL, Brown BW, Bandura A, Ritter P, et al. Evidence suggesting that a chronic disease self-management program can improve health status while reducing hospitalization: a randomized trial. Med Care 1999;37(1):5-14. [doi: 10.1097/00005650-199901000-00003] [Medline: 10413387]
20. Perry J, VanDenKerkhof EG, Wilson R, Tripp DA. Development of a guided internet-based psycho-education intervention using cognitive behavioral therapy and self-management for individuals with chronic pain. Pain Manag Nurs 2017;18(2):90-101. [doi: 10.1016/j.pmn.2016.12.004] [Medline: 28262481]
21. Ruehlman LS, Karoly P, Enders C. A randomized controlled evaluation of an online chronic pain self management program. Pain 2012;153(2):319-330 [FREE Full text] [doi: 10.1016/j.pain.2011.10.025] [Medline: 22133450]
22. Stinson JN, Lalloo C, Harris L, Isaac L, Campbell F, Brown S, et al. iCanCope with Pain User-centred design of a web- and mobile-based self-management program for youth with chronic pain based on identified health care needs. Pain Res Manag 2014;19(5):257-265 [FREE Full text] [doi: 10.1155/2014/935278] [Medline: 25000507]
23. Tse M, Li Y, Tang SK, Ng SSM, Bai X, Lee P, et al. An exploration of the effectiveness of a peer-led pain management program (PAP) for nursing home residents with chronic pain and an evaluation of their experiences: a pilot randomized controlled trial. Int J Environ Res Public Health 2020;17(11):4090 [FREE Full text] [doi: 10.3390/ijerph17114090] [Medline: 32521785]




XSL•FO
RenderX

24. Yuan SLK, Marques AP. Development of ProFibro-a mobile application to promote self-care in patients with fibromyalgia. Physiotherapy 2018;104(3):311-317. [doi: 10.1016/j.physio.2018.04.005] [Medline: 30025715]
25. Trindade IA, Guiomar R, Carvalho SA, Duarte J, Lapa T, Menezes P, et al. Efficacy of online-based acceptance and commitment therapy for chronic pain: a systematic review and meta-analysis. J Pain 2021;22(11):1328-1342 [FREE Full text] [doi: 10.1016/j.jpain.2021.04.003] [Medline: 33892153]
26. de Lorena SB, Duarte ALBP, Bredemeier M, Fernandes VM, Pimentel EAS, Marques CDL, et al. Effects of a physical self-care support program for patients with fibromyalgia: a randomized controlled trial. J Back Musculoskelet Rehabil 2022;35(3):495-504. [doi: 10.3233/BMR-191820] [Medline: 34657869]
27. Musekamp G, Gerlich C, Ehlebracht-Kï Nig I, Dorn M, Hï Fter A, Tomiak C, et al. Evaluation of a self-management patient education programme for fibromyalgia-results of a cluster-RCT in inpatient rehabilitation. Health Educ Res 2019;34(2):209-222. [doi: 10.1093/her/cyy055] [Medline: 30689860]
28. Ledel Solem IK, Varsi C, Eide H, Kristjansdottir OB, Børøsund E, Schreurs KMG, et al. A user-centered approach to an evidence-based electronic health pain management intervention for people with chronic pain: design and development of EPIO. J Med Internet Res 2020;22(1):e15889 [FREE Full text] [doi: 10.2196/15889] [Medline: 31961331]
29. Borrelli B, Ritterband LM. Special issue on eHealth and mHealth: challenges and future directions for assessment, treatment, and dissemination. Health Psychol 2015;34S:1205-1208. [doi: 10.1037/hea0000323] [Medline: 26651461]
30. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3(1):4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
31. Krusche M, Ruffer N, Grahammer M, Knitza J. Mobile apps and their usage in rheumatology. Z Rheumatol 2020;79(6):554-561. [doi: 10.1007/s00393-020-00822-2] [Medline: 32472178]
32. Shetty A, Delanerolle G, Zeng Y, Shi JQ, Ebrahim R, Pang J, et al. A systematic review and meta-analysis of digital application use in clinical research in pain medicine. Front Digit Health 2022:4 [FREE Full text] [doi: 10.1101/2022.03.02.22271773]
33. Valentijn PP, Tymchenko L, Jacobson T, Kromann J, Biermann CW, AlMoslemany MA, et al. Digital health interventions for musculoskeletal pain conditions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022;24(9):e37869 [FREE Full text] [doi: 10.2196/37869] [Medline: 36066943]
34. Pfeifer AC, Uddin R, Schröder-Pfeifer P, Holl F, Swoboda W, Schiltenwolf M. Mobile application-based interventions for chronic pain patients: a systematic review and meta-analysis of effectiveness. J Clin Med 2020;9(11):3557 [FREE Full text] [doi: 10.3390/jcm9113557] [Medline: 33167300]
35. Ehde DM, Dillworth TM, Turner JA. Cognitive-behavioral therapy for individuals with chronic pain: efficacy, innovations, and directions for research. Am Psychol 2014;69(2):153-166. [doi: 10.1037/a0035747] [Medline: 24547801]
36. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
37. Goldenberg DL, Burckhardt C, Crofford L. Management of fibromyalgia syndrome. JAMA 2004;292(19):2388-2395. [doi: 10.1001/jama.292.19.2388] [Medline: 15547167]
38. Winkelmann A, Häuser W, Friedel E, Moog-Egan M, Seeger D, Settan M, et al. Physiotherapy and physical therapies for fibromyalgia syndrome. Systematic review, meta-analysis and guideline. Schmerz 2012;26(3):276-286. [doi: 10.1007/s00482-012-1171-3] [Medline: 22760460]
39. Köllner V, Häuser W, Klimczyk K, Kühn-Becker H, Settan M, Weigl M, et al. Psychotherapy for patients with fibromyalgia syndrome. Systematic review, meta-analysis and guideline. Schmerz 2012;26(3):291-296. [doi: 10.1007/s00482-012-1179-8] [Medline: 22760462]
40. Langhorst J, Häuser W, Bernardy K, Lucius H, Settan M, Winkelmann A, et al. Complementary and alternative therapies for fibromyalgia syndrome. Systematic review, meta-analysis and guideline. Schmerz 2012;26(3):311-317. [doi: 10.1007/s00482-012-1178-9] [Medline: 22760464]
41. Albuquerque MLL, Monteiro D, Marinho DA, Vilarino GT, Andrade A, Neiva HP. Effects of different protocols of physical exercise on fibromyalgia syndrome treatment: systematic review and meta-analysis of randomized controlled trials. Rheumatol Int 2022;42(11):1893-1908. [doi: 10.1007/s00296-022-05140-1] [Medline: 35604435]
42. Bidonde J, Busch AJ, Schachter CL, Overend TJ, Kim SY, Góes SM, et al. Aerobic exercise training for adults with fibromyalgia. Cochrane Database Syst Rev 2017(6):CD012700. [doi: 10.1002/14651858.cd012700]
43. Bidonde J, Busch AJ, Schachter CL, Webber SC, Musselman KE, Overend TJ, et al. Mixed exercise training for adults with fibromyalgia. Cochrane Database Syst Rev 2019;5(5):CD013340 [FREE Full text] [doi: 10.1002/14651858.CD013340] [Medline: 31124142]
44. Busch AJ, Webber SC, Richards RS, Bidonde J, Schachter CL, Schafer LA, et al. Resistance exercise training for fibromyalgia. Cochrane Database Syst Rev 2013;2013(12):CD010884 [FREE Full text] [doi: 10.1002/14651858.CD010884] [Medline: 24362925]
45. Busch AJ, Barber KAR, Overend TJ, Peloso PMJ, Schachter CL. Exercise for treating fibromyalgia syndrome. Cochrane Database Syst Rev 2007(4). [doi: 10.1002/14651858.cd003786]




XSL•FO
RenderX

46. Couto N, Monteiro D, Cid L, Bento T. Effect of different types of exercise in adult subjects with fibromyalgia: a systematic review and meta-analysis of randomised clinical trials. Sci Rep 2022;12(1):10391 [FREE Full text] [doi: 10.1038/s41598-022-14213-x] [Medline: 35725780]
47. da Silva JM, de Barros BS, Almeida GJ, O'Neil J, Imoto AM. Dosage of resistance exercises in fibromyalgia: evidence synthesis for a systematic literature review up-date and meta-analysis. Rheumatol Int 2022;42(3):413-429. [doi: 10.1007/s00296-021-05025-9] [Medline: 34652480]
48. Kim SY, Busch AJ, Overend TJ, Schachter CL, van der Spuy I, Boden C, et al. Flexibility exercise training for adults with fibromyalgia. Cochrane Database Syst Rev 2019;9(9):CD013419 [FREE Full text] [doi: 10.1002/14651858.CD013419] [Medline: 31476271]
49. Bernardy K, Klose P, Busch AJ, Choy EHS, Häuser W. Cognitive behavioural therapies for fibromyalgia. Cochrane Database Syst Rev 2013;2013(9):CD009796 [FREE Full text] [doi: 10.1002/14651858.CD009796.pub2] [Medline: 24018611]
50. de Oliveira Lima L, Saragiotto B, Costa L, Nogueira L, Meziat-Filho N, Reis F. Self-guided web-based pain education for people with musculoskeletal pain: a systematic review and meta-analysis. Phys Ther 2021;101(10):pzab167. [doi: 10.1093/ptj/pzab167] [Medline: 34174081]
51. Du S, Dong J, Jin S, Zhang H, Zhang Y. Acceptance and commitment therapy for chronic pain on functioning: a systematic review of randomized controlled trials. Neurosci Biobehav Rev 2021;131:59-76. [doi: 10.1016/j.neubiorev.2021.09.022] [Medline: 34536462]
52. Enomoto K, Adachi T, Fujino H, Kugo M, Tatsumi S, Sasaki J. Comparison of the effectiveness of cognitive behavioral therapy for insomnia, cognitive behavioral therapy for pain, and hybrid cognitive behavioral therapy for insomnia and pain in individuals with comorbid insomnia and chronic pain: a systematic review and network meta-analysis. Sleep Med Rev 2022;66:101693. [doi: 10.1016/j.smrv.2022.101693] [Medline: 36410076]
53. Gandy M, Pang STY, Scott AJ, Heriseanu AI, Bisby MA, Dudeney J, et al. Internet-delivered cognitive and behavioural based interventions for adults with chronic pain: a systematic review and meta-analysis of randomized controlled trials. Pain 2022;163(10):e1041-e1053. [doi: 10.1097/j.pain.0000000000002606] [Medline: 35121696]
54. García-Ríos MC, Navarro-Ledesma S, Tapia-Haro RM, Toledano-Moreno S, Casas-Barragán A, Correa-Rodríguez M, et al. Effectiveness of health education in patients with fibromyalgia: a systematic review. Eur J Phys Rehabil Med 2019;55(2):301-313 [FREE Full text] [doi: 10.23736/S1973-9087.19.05524-2] [Medline: 30698402]
55. Joypaul S, Kelly F, McMillan SS, King MA. Multi-disciplinary interventions for chronic pain involving education: a systematic review. PLoS One 2019;14(10):e0223306 [FREE Full text] [doi: 10.1371/journal.pone.0223306] [Medline: 31577827]
56. Louw A, Zimney K, Puentedura EJ, Diener I. The efficacy of pain neuroscience education on musculoskeletal pain: a systematic review of the literature. Physiother Theory Pract 2016;32(5):332-355. [doi: 10.1080/09593985.2016.1194646] [Medline: 27351541]
57. Ma TW, Yuen ASK, Yang Z. The efficacy of acceptance and commitment therapy for chronic pain: a systematic review and meta-analysis. Clin J Pain 2023;39(3):147. [doi: 10.1097/ajp.0000000000001096]
58. Romm MJ, Ahn S, Fiebert I, Cahalin LP. A meta-analysis of therapeutic pain neuroscience education, using dosage and treatment format as moderator variables. Pain Pract 2021;21(3):366-380. [doi: 10.1111/papr.12962] [Medline: 33131210]
59. Veehof MM, Trompetter HR, Bohlmeijer ET, Schreurs KMG. Acceptance- and mindfulness-based interventions for the treatment of chronic pain: a meta-analytic review. Cogn Behav Ther 2016;45(1):5-31 [FREE Full text] [doi: 10.1080/16506073.2015.1098724] [Medline: 26818413]
60. Williams ACC, Fisher E, Hearn L, Eccleston C. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2020;8(8):CD007407 [FREE Full text] [doi: 10.1002/14651858.CD007407.pub4] [Medline: 32794606]
61. Cheng CA, Chiu YW, Wu D, Kuan YC, Chen SN, Tam KW. Effectiveness of Tai Chi on fibromyalgia patients: a meta-analysis of randomized controlled trials. Complement Ther Med 2019;46:1-8. [doi: 10.1016/j.ctim.2019.07.007] [Medline: 31519264]
62. Langhorst J, Klose P, Dobos GJ, Bernardy K, Häuser W. Efficacy and safety of meditative movement therapies in fibromyalgia syndrome: a systematic review and meta-analysis of randomized controlled trials. Rheumatol Int 2013;33(1):193-207. [doi: 10.1007/s00296-012-2360-1] [Medline: 22350253]
63. Lee C, Crawford C, Schoomaker E. Movement therapies for the self-management of chronic pain symptoms. Pain Med 2014;15(S1):S40-S53. [doi: 10.1111/pme.12411] [Medline: 24734859]
64. Lee C, Crawford C, Hickey A. Mind-body therapies for the self-management of chronic pain symptoms. Pain Med 2014;15(S1):S21-S39. [doi: 10.1111/pme.12383] [Medline: 24734857]
65. Theadom A, Cropley M, Smith HE, Feigin VL, McPherson K. Mind and body therapy for fibromyalgia. Cochrane Database Syst Rev 2015;2015(4):CD001980 [FREE Full text] [doi: 10.1002/14651858.CD001980.pub3] [Medline: 25856658]
66. Hernando-Garijo I, Jiménez-del-Barrio S, Mingo-Gómez T, Medrano-de-la-Fuente R, Ceballos-Laita L. Effectiveness of non-pharmacological conservative therapies in adults with fibromyalgia: a systematic review of high-quality clinical trials. J Back Musculoskelet Rehabil IOS Press 2022;35(1):3-20. [doi: 10.3233/bmr-200282]




XSL•FO
RenderX

67. Kundakci B, Kaur J, Goh SL, Hall M, Doherty M, Zhang W, et al. Efficacy of nonpharmacological interventions for individual features of fibromyalgia: a systematic review and meta-analysis of randomised controlled trials. Pain 2022;163(8):1432-1445. [doi: 10.1097/j.pain.0000000000002500] [Medline: 34813518]
68. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ British Medical Journal 2000;321(7262):694-696 [FREE Full text] [doi: 10.1136/bmj.321.7262.694] [Medline: 10987780]
69. American COSM, Liguori G, Feito Y, Fountaine C, Roy B. ACSM's Guidelines for Exercise Testing and Prescription. 11th ed. Philadelphia: Wolters Kluwer; 2021.
70. Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J. Acceptance and commitment therapy: model, processes and outcomes. Behav Res Ther 2006;44(1):1-25. [doi: 10.1016/j.brat.2005.06.006] [Medline: 16300724]
71. Moseley GL, Butler DS. Fifteen years of explaining pain: the past, present, and future. J Pain 2015;16(9):807-813 [FREE Full text] [doi: 10.1016/j.jpain.2015.05.005] [Medline: 26051220]
72. National Library of Medicine. NCBI Mesh. NCBI NLM NIH. URL: https://www-ncbi-nlm-nih-gov.ezproxy.is.ed.ac.uk/mesh/ [accessed 2023-06-24]
73. Falci L, Shi Z, Greenlee H. Multiple chronic conditions and use of complementary and alternative medicine among US adults: results from the 2012 national health interview survey. Prev Chronic Dis 2016;13:E61 [FREE Full text] [doi: 10.5888/pcd13.150501] [Medline: 27149072]
74. Miles CL, Pincus T, Carnes D, Homer KE, Taylor SJC, Bremner SA, et al. Can we identify how programmes aimed at promoting self-management in musculoskeletal pain work and who benefits? A systematic review of sub-group analysis within RCTs. Eur J Pain 2011;15(8):775.e1-775.11. [doi: 10.1016/j.ejpain.2011.01.016] [Medline: 21354838]
75. Chew MT, Chan C, Kobayashi S, Cheng HY, Wong TM, Nicholson LL. Online pain management programs for chronic, widespread musculoskeletal conditions: a systematic review with meta-analysis. Pain Pract 2023;23(6):664-683. [doi: 10.1111/papr.13227] [Medline: 37051894]
76. DIGAs. Bundesinstitut für Arzneimittel und Medizinprodukte. URL: https://diga.bfarm.de/de [accessed 2024-06-16]
77. Gendreau RM, McCracken LM, Williams DA, Luciano JV, Dai Y, Vega N, et al. Self-guided digital behavioural therapy versus active control for fibromyalgia (PROSPER-FM): a phase 3, multicentre, randomised controlled trial. Lancet 2024;404(10450):364-374. [doi: 10.1016/S0140-6736(24)00909-7] [Medline: 38991582]
78. Blanchard M, Koller CN, Azevedo PM, Prétat T, Hügle T. Development of a management app for postviral fibromyalgia-like symptoms: patient preference-guided approach. JMIR Form Res 2024;8:e50832 [FREE Full text] [doi: 10.2196/50832] [Medline: 38639986]
79. Blanchard M, Venerito V, Ming Azevedo P, Hügle T. Generative AI-based knowledge graphs for the illustration and development of mHealth self-management content. Front Digit Health 2024;6:1466211 [FREE Full text] [doi: 10.3389/fdgth.2024.1466211] [Medline: 39434919]
80. Remedee Labs. URL: https://remedeelabs.com/en/ [accessed 2025-12-13]
81. Schmajuk G, Nasrallah C, Berrean B, Prugh J, Wilson C, Hamblin A, et al. A step-by-step roadmap for the development and deployment of an electronic health record sidecar application that tracks patient outcomes: the RA PRO dashboard. Digit Health 2024;10:20552076241288739 [FREE Full text] [doi: 10.1177/20552076241288739] [Medline: 39421306]


## Abbreviations

**ACT:** acceptance and commitment therapy
**AI:** artificial intelligence
**CBT:** cognitive behavioral therapy
**DiGA:** Digitale Gesundheitsanwendung
**FIQ:** Fibromyalgia Impact Questionnaire
**LLM:** large language model
**mHealth:** mobile health
**MMT:** meditative movement therapy
**QoL:** quality of life



XSL•FO
RenderX

*Edited by N Cahill; submitted 16.Oct.2024; peer-reviewed by A D'Silva, Q Ye; comments to author 23.Dec.2024; accepted 16.Nov.2025; published 17.Feb.2026.*

*Please cite as:*
*Lebeau Foustoukos T, Leclercq I, Blanchard M, Hügle T*
*Evidence-Based Self-Management Strategies for Fibromyalgia: Foundations for Digital Therapeutic Applications*
*Interact J Med Res 2026;15:e67523*
*URL: https://www.i-jmr.org/2026/1/e67523*
*doi:10.2196/67523*
*PMID:*

©Thomas Lebeau Foustoukos, Isabelle Leclercq, Marc Blanchard, Thomas Hügle. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 17.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# How Many Instruments Do We Need? Strengths, Weaknesses, and Potentials for Further Developments of Digital Health Literacy Measurement Instruments: Narrative Overview of Reviews and Content Analysis

Pauline Kaboth, MSc; Lorenz Harst, MA, Dr rer medic; Jochen Schmitt, MPH, Prof Dr med; Madlen Scheibe, Dipl Soz, Dr rer medic

Center for Evidence-Based Healthcare, Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, Fetscherstraße 74, Dresden, Germany

**Corresponding Author:**
Pauline Kaboth, MSc
Center for Evidence-Based Healthcare, Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, Fetscherstraße 74, Dresden, Germany

## Abstract

**Background:** With the ongoing digitalization of health care, digital health literacy (DHL) is becoming increasingly important, requiring appropriate measurement instruments (DHLMIs). However, the wide range of available DHLMIs makes selection difficult and raises questions about their suitability and comparability.

**Objective:** This review aimed to provide an overview of available DHLMIs for adult populations and to compare their dimensions, identifying overlaps and differences to determine which (key) dimensions are most commonly studied and thus define DHL.

**Methods:** A narrative overview of reviews was conducted. The database search was conducted in November 2024, with an updated search in March 2026, in PubMed and Google Scholar using specific search terms and predefined inclusion and exclusion criteria. This was complemented by a forward citation search in Web of Science. All identified records were screened in a multistage process. At the review level, systematic and scoping reviews published since January 1, 2020, were included that analyzed DHLMIs in adult populations. At the primary study level, studies were included in which DHLMIs were used for self-assessment, performance-based evaluation, or a combination of both. Data extraction was performed by one reviewer and verified by a second reviewer. Data on underlying theories, methods of data collection (performance-based or self-reported), and target groups were extracted. DHL dimensions, their definitions, and the associated items were categorized by 2 researchers using qualitative content analysis according to Kuckartz.

**Results:** A total of 12 reviews were included. Of the 170 measurement instruments examined in these reviews, 33 (19%) were selected for detailed analysis after applying the inclusion and exclusion criteria. The majority of the included measurement instruments (n=20, 61%) were not based on a theory, 20 (61%) measured DHL exclusively via self-report, and 21 (64%) addressed specific target groups. The measurement instruments encompassed a total of 209 original dimensions of DHL. The number of dimensions measured per instrument varied between 2 and 22. The qualitative content analysis identified a total of 30 assigned dimensions. Key dimensions captured in almost all instruments, although under different names, include evaluating health information, using health information, researching health information, and the ability to use technology.

**Conclusions:** The large number of measurement instruments and original dimensions makes it difficult to select suitable instruments for measuring DHL and to compare and synthesize study results. Nevertheless, we were able to identify key dimensions as relevant regardless of target groups and types of digital applications, which could be used to develop a core outcome set for DHL. Our findings offer a foundation for refining existing instruments and developing new ones. They provide practical guidance for researchers and health care professionals in selecting suitable DHLMIs. Additionally, our review underlines the importance of theory to ensure content validity and comparability of DHLMIs.

***(Interact J Med Res 2026;15:e91603)*** doi:10.2196/91603

**KEYWORDS**

electronic health literacy; eHealth literacy; digital health; telemedicine; review; qualitative research; digital health literacy



XSL•FO
RenderX

## Introduction

### Background

Health literacy is a key resource for achieving and maintaining good health [1,2]. It is defined as the knowledge, motivation, and ability of individuals to access, understand, evaluate, and apply health information to make informed decisions that support or improve health and quality of life [3]. While health information was traditionally accessed through print media, (audio-)visual broadcasts, or direct interaction with health care professionals, due to increasing digitalization, the internet has become the dominant source of health-related information [4,5].

To enable effective use of this information, individuals need to have digital health literacy (DHL). Understanding how this concept has developed over time is essential for contextualizing the diversity of available measurement instruments and the dimensions they cover. The concept of eHealth literacy, on which DHL is based, was first defined by Norman and Skinner in 2006 as the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to address or solve a health problem [6,7]. Their Lily Model described eHealth literacy as a set of 6 core literacies: traditional literacy, health literacy, information literacy, scientific literacy, media literacy, and computer literacy [7]. As digital health technologies evolved beyond static web resources to include interactive, mobile, and social media platforms, the concept of "digital health literacy" was introduced to capture the broader and more complex skill set required to effectively engage with these new environments [8]. Norman [9] himself acknowledged these limitations in his 2011 update, and subsequent scholars have proposed revised or entirely new definitions and models (eg, [10,11]). This progress has prompted the development of broader conceptual frameworks. The conceptual heterogeneity of the field is illustrated by a concept analysis by Ban et al [8], which identified multiple competing definitions and traced how the concept has continued to evolve in response to technological and contextual change.

The diversity of available theoretical models, coupled with the absence of a universally accepted definition, is reflected in the large number of measurement instruments available [10-12]. These differ in their underlying theoretical frameworks, target populations, assessment methods—ranging from performance-based to self-reported approaches—and the dimensions they cover. The wide range of instruments available makes selection difficult and raises questions about the suitability and comparability of study results. This challenge is not limited to DHL but is also present in clinical settings, for example, in the assessment of disease severity using instruments for atopic dermatitis [13].

Previous reviews have analyzed the psychometric properties of available digital health measurement instruments (DHLMIs) using the COSMIN methodology [12,14] or conducted content analyses restricted to instrument development studies for the general population [15] or focused exclusively on either self-reported [12] or performance-based instruments [16]. However, no review has yet applied a structured qualitative content analysis to systematically examine and synthesize the dimensional landscape across the full breadth of DHLMIs in active research use, including instruments for both the general population and specific target groups, such as older adults or individuals with chronic conditions. This review aimed to address this gap.

### Study Objectives

The objective of our study was to provide a comprehensive overview of existing DHLMIs, both self-reported and performance-based, to systematically analyze and synthesize the dimensions across instruments, including those developed for the general population and those targeting specific populations, such as older adults or individuals with chronic conditions. We focused on the underlying theory, assessed dimensions, assessment methods (ie, performance-based, self-report, or mixed method approaches), and target populations. Rather than aiming to identify a single universal instrument, our goal was to identify overlaps and differences between instruments and to determine which key dimensions are most commonly operationalized across DHLMIs, thereby contributing to a more systematic understanding of how digital health literacy is currently conceptualized and measured in research practice. This analysis is intended to support researchers in making more informed decisions about which instrument to use, to provide instrument developers with an evidence-based starting point, and to contribute to the development of a core outcome set for DHL.

## Methods

### Overview

We conducted a narrative overview of reviews, including their primary studies, to provide a meta-level synthesis concerning the aforementioned research aim. We opted for a narrative review because it offers a more flexible structure and is not as strictly bound to a search and evaluation protocol. This allows us to more easily integrate different perspectives, theories, and study types into this study and provide a broader overview of the research field [17]. This approach is particularly suitable for heterogeneous topics where existing research is diverse, making it easier to present and discuss the variety of approaches and findings [18].

Another reason for choosing a narrative review is that it supports our primary objective. Our primary objective was to analyze the dimensions represented in the measurement instruments. This question is inherently qualitative in nature and is best addressed through an interpretive synthesis. Despite our narrative approach, we have applied several systematic principles in key areas of our work, which are described in more detail below.

### Search Strategy

We conducted an electronic database search on November 4, 2024, with an additional update search on March 26, 2026. Searches were performed in PubMed and Google Scholar to capture relevant interdisciplinary sources. Furthermore, a forward citation search in Web of Science was conducted as part of the update search to enhance the comprehensiveness of the literature search.



XSL•FO
RenderX

Our search terms included the free text terms: “Digital Health Literacy,” “eHealth Literacy,” “Digital Literacy,” “Digital Health Competence,” “Health Information Literacy,” “Digital Health Skills,” “Digital Health Knowledge,” “Online Health Literacy,” and “Electronic Health Literacy,” in various combinations with the terms “tool” or “instrument” via the Boolean operator “AND.”

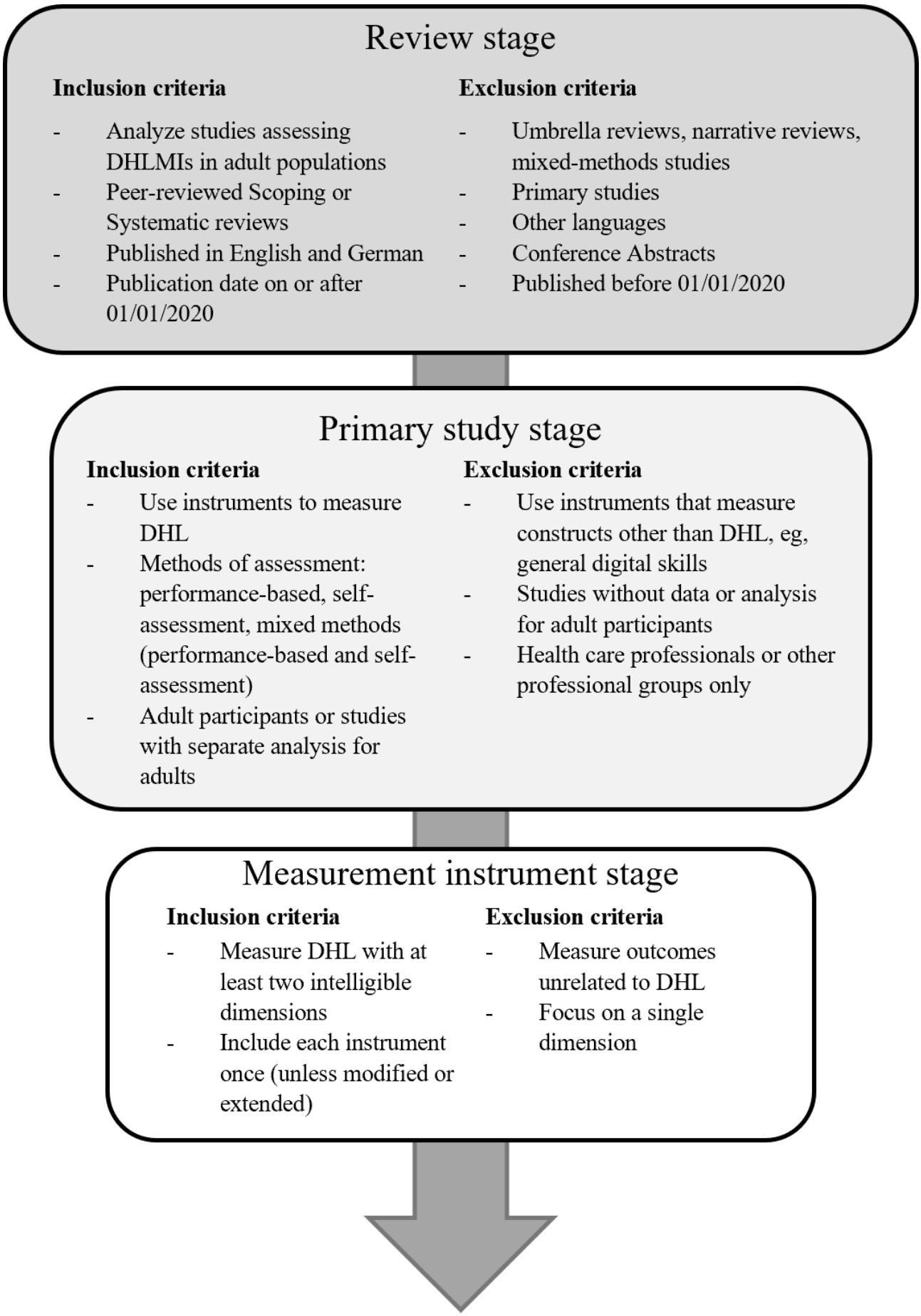
### Study Selection

A 3-stage inclusion process at the review, primary study, and measurement instrument level was used to arrive at the number of instruments relevant to our analysis (Figure 1).



XSL•FO
RenderX

**Figure 1.** Inclusion and exclusion criteria. DHL: digital health literacy; DHLMI: digital health measurement instruments.






XSL•FO
RenderX

At the review stage, we included reviews that analyzed studies assessing DHLMIs in adult populations. Specifically, we considered systematic and scoping reviews published in English or German since January 1, 2020, to present the current evidence. In addition, the search periods of the reviews also cover primary studies from earlier years, which, in turn, examine measurement instruments developed at that time. We excluded umbrella reviews, mixed methods studies, narrative reviews, primary studies, and research published in languages other than English or German. Conference abstracts and studies published before January 1, 2020, were also excluded. To focus on the most recent evidence on the topic, we also excluded conference abstracts and studies published before January 1, 2020. At the primary study stage, we included studies that used instruments to measure DHL through self-report methods, performance-based assessments, or a combination of both. We restricted inclusion to studies conducted with adult participants or studies that included mixed age groups if a separate analysis for adult participants was available. Studies conducted exclusively with health care professionals or other professional groups were excluded, as this review focuses on DHLMIs designed to assess personal DHL in the context of individual health management, rather than professional digital health competence applied in clinical practice.

At the instrument stage, we included instruments published in English or German that measured DHL by considering at least two intelligible dimensions. DHL is conceptualized as an inherently multidimensional construct across established theoretical frameworks, including the Lily Model [7], the eHealth Literacy Framework [19], and the Transactional Model of eHealth Literacy [20]. Instruments capturing only a single dimension fail to adequately operationalize this complexity. Therefore, we restricted inclusion to instruments measuring at least two distinct dimensions of DHL. Additionally, we excluded instruments that measured only health literacy without digital attributes or digital literacy without health-related components.

If one measurement instrument was used in more than one study, we only included the instrument once in our analyses, unless the original measurement instrument was supplemented or modified in another study. If a measurement instrument was used in multiple studies, we selected the most recent primary study to ensure that all previous developments and refinements of the instrument were taken into account.

### Data Extraction

We developed a standardized data extraction sheet within the research team. All relevant information on the included reviews, primary studies, and measurement instruments—including dimensions, theory, methods of assessment (performance-based or self-assessment), target groups, contexts of use, and purpose—was extracted and recorded in said extraction sheet. Data extraction was carried out by one reviewer (PK) and verified for accuracy and completeness by a second reviewer (LH).

### Data Synthesis

To determine which conceptual dimensions (*original dimensions*) were covered by the included measurement instruments, we conducted a structured qualitative content analysis following the approach by Kuckartz [21]. Therefore, we reviewed the dimension definitions and items from each instrument and grouped them into broader categories, which we called *assigned dimensions*. Where dimensions were not explicitly described, the assignment to our categories was based directly on the items. These categories were developed and refined through joint coding by 2 researchers (PK and LH). Differences in coding were discussed and resolved by consensus within the entire research team (PK, LH, and MS). Dimensions for which no clear definitions or corresponding items were available or accessible were excluded from this analysis. Subsequently, the assigned dimensions were systematically grouped into overarching domains.

To assess the prevalence of each assigned dimension, we applied two counts of frequency: (1) frequency across instruments, where a dimension was counted once per instrument regardless of how often it appeared within it; and (2) frequency per original dimension, where each occurrence of the assigned dimension across different original dimensions within the same instrument was counted separately. The latter step was necessary as we found different definitions for the measurement instruments for the assigned dimensions. This dual approach enabled us to capture both the overall representation and the specific distribution of each assigned dimension.

## Results

### Overview of Included Studies Focusing on DHLMIs

We identified 12 reviews [4,11,12,14-16,22-27] that met the inclusion criteria and were included in our analysis (Table 1).

These reviews were published between 2021 and 2026 and included a total of 589 primary studies. Five (42%) of the included reviews were systematic reviews [4,12,14,24,27], and the remaining reviews were scoping reviews. Seven (58%) reviews examined DHLMIs for specific target groups or settings such as adults living in community environments [4], people in hospital settings [23], older adults [22,24,26,27], or patients of health and social care professionals [25], while the other reviews studied DHL as a generic construct [11,12,14-16]. Four (33%) reviews [4,11,12,14] considered only self-reported measurements, and one (8%) review [16] examined only performance-based measurements. Six (50%) of the 12 reviews investigated both self-report and performance-based measurements [15,23-27]. The median number of primary studies included was 40 (range 2-251). The median number of included instruments was 10, with a range of 2 to 44. The primary studies included in the reviews comprised a total of 170 instruments. After removing the duplicates, a total of 94 instruments remained. Of these 94 instruments, 51 measured DHL, of which 33 met our inclusion and exclusion criteria.



XSL•FO
RenderX

**Table .** Characteristics of included reviews.

| Author, year | Study design | Inclusion period (as reported) | Aim of the study regarding DHL[a] instruments | Performance-based or self-reported | Included studies, n | Participants in all included primary studies, N | Included instruments, n | Included instruments measuring DHL, n | Target group or setting of the review |
|---|---|---|---|---|---|---|---|---|---|
| Ahn and Kim, 2024 [4] | Systematic review | NR[b] | Evaluate the characteristics and psychometric properties of questionnaires designed to measure DHL | S[c] | 21 | 11,325 | 21 | 19 | Adolescents and adults living in community settings |
| Bai et al, 2025 [22] | Scoping review | 2014 and March 2024 | Identify tools for assessing digital health literacy to guide dietetic practice in screening patients for digital care needs and training | S | 66 | 40,517 | 8 | 8 | Middle-aged and older adults (age >45 y) in clinical, community, or population settings |
| Crocker et al, 2023 [16] | Scoping review | Up to June 2021 | Identify tools measuring DHL based on objective performance | P[d] | 33 | 8404 | 33 | 29 | No specific target group |
| Délétroz et al, 2026 [14] | Systematic review | January 2000 to 2024 | Evaluate measurement properties of patient-reported outcome measures of eHealth literacy | S | 89 studies and 3 reports | NR | 17 | 15 | Adult population (age >18 y) |
| Dijkman et al, 2023 [23] | Scoping review | 1991 - 2022 | Provide an overview of tools for measuring (digital) health literacy in hospitals | P and S | 251 | NR | 44 | 7 | Hospital setting |
| Faux-Nightingale et al, 2022 [11] | Scoping review | Up to February 25, 2021 | Assess available tools that can be used to evaluate DHL | S | 56 | NR | 5 | 5 | No specific target group/setting |
| Huang et al, 2023 [24] | Systematic review | Up to January 13, 2021 | Examine the diagnostic accuracy of DHL tools | P and S | 2 | 365 | 2 | 2 | Older adults (age >60 y) |



XSL•FO
RenderX

| Author, year | Study design | Inclusion period (as reported) | Aim of the study regarding DHL[a] instruments | Performance-based or self-reported | Included studies, n | Participants in all included primary studies, N | Included instruments, n | Included instruments measuring DHL, n | Target group or setting of the review |
|---|---|---|---|---|---|---|---|---|---|
| Kaihlanen et al, 2023 [25] | Scoping review | Up to February 2023 | Identify available evaluation tools for assessing the client's potential and suitability for using digital health | P and S | 19 | NR | 12 | 9 | Patients of health and social care professionals |
| Lee et al, 2021 [12] | Systematic review | Up to March 3, 2021 | Identify the currently available instruments for measuring eHealth literacy and to evaluate their measurement properties | S | 57 | NR | 7 | 7 | No specific target group or setting |
| Wang and Luan, 2022 [26] | Scoping review | NR | Identify available evaluation tools to assess DHL | P and S | 47 | NR | 5 | 5 | Older adults (age ≥65 y) |
| Wang et al, 2025 [15] | Scoping review | 2006 to June 2024 | Evaluate the characteristics, effectiveness, and limitations of existing eHealth literacy assessment instruments | P and S | 13 | NR | 13 | 13 | No specific target group |
| Xie et al, 2022 [27] | Systematic review | NR | Identify, appraise, and synthesize research evidence of the association between DHL and health outcomes | P and S | 24 | 11,778 | 3 | 3 | Older adults (age >60 y) |

[a]DHL: digital health literacy.

[b]NR: not reported in the reviews.

[c]S: self-reported.

[d]P: performance based.

## Overview of DHLMIs

In total, 33 different DHLMIs were identified that met our inclusion criteria (Table 2). These were used 234 times in the primary studies (Table 3). The measurement instruments reviewed by Délétroz et al [14] and Wang et al [15,26] are not included in this count, as none of these reviews reported information on the frequency of use of instruments in primary studies. The most frequently used instruments used in the primary studies included in the reviews were the 8-item eHealth Literacy Scale [6] (n=163, 72%), the eHealth Literacy Questionnaire [28] (n=13, 6%), the Digital Health Literacy Instrument Screening Tool [29] (n=12, 5%), the eHealth Literacy Assessment toolkit [30] (n=8, 3%), the Transactional eHealth Literacy Instrument [20] (n=5, 2%), the Electronic-Health Literacy Scale [31] (n=3, 1%), and the Extended eHealth



XSL•FO
RenderX

Literacy Scale [32] (n=3, 1%). Three instruments (Digital Health Technology Literacy Assessment Questionnaire [33], eHEALS+not specified [34], and Readiness and Enablement Index for Health Technology [35]) were used twice. All other 23 instruments were used in only one primary study [29,36-57].

These were developed between 2006 and 2023. Thirteen (40%) of these were developed in North America [6,20,30,31,37,38,40-42,44,50,53,56], 10 (30%) were developed in Asia [33,34,36,39,43,45-47,51,57], and 10 (30%) were developed in Europe [28,29,32,35,48,49,52,54,55,58]. For most instruments (n=20, 61%), no theoretical basis for the development of the instruments was reported (Table 2) [29,31,35,37,39-41,43,45-48,50,52-58]. The Lily Model served as the theoretical basis for 8 (24%) instruments [6,30,32,34,36,42,44,49]. The remaining instruments not developed from scratch without any theoretical foundation were based on other theories, such as the Transactional Model of eHealth Literacy (TMeHL) [20].

Across all instruments, a total of 209 original dimensions of DHL were identified (Multimedia Appendix 1).

The number of original dimensions per instrument ranged from 2 [33,43] to 22 [53]. The sample size of the studies on instrument development ranged from 15 [29] to 1914 participants [34]. Of these instruments, 12 (36%) were designed for the general population [6,28,30,31,37,47,48,51,54,55,57,58], and 21 (64%) targeted specific population groups [20,29,32-36,38-46,49,50,52,53,56], such as students or older people with chronic diseases. The average age of participants in these studies ranged from 12.9 (SD 0.9) [42] to 71.27 (SD 10.82) years [45]. In terms of gender, 14 (43%) studies had an almost equal proportion of male and female participants [20,30-34,36,44-46,51,54,55,57,58]. In the remaining studies, the study population was predominantly women (n=12, 36%) [29,35,37,40-42,47,49,50,53,56] and men (n=4, 12%) [6,43,52,54]. Three (9%) studies [28,39,48] did not report on the gender distribution in the analyzed population.

The DHLMIs varied in their design: 7 (21%) of the instruments measured DHL on a performance basis [29,37,38,40,47,53,55], 20 (61%) applied self-reported measurements [6,20,28,30-33,35,39,42-46,48-52,57], and 6 (18%) combined both approaches [29,34,36,41,54,56].



XSL•FO
RenderX

**Table .** Characteristics of included measurement instruments.

| Instrument(s) | Author, year | Theory | Country | Participants included in primary studies, N | Generic instrument or for specific target group | Age (y), mean (SD) or range | Women (%) | Performance based or self-reported |
|---|---|---|---|---|---|---|---|---|
| CeHLS-D[a] | Lee et al, 2022 [43] | None specified | South Korea | 453 | T[b] (patients with chronic diseases) | 56.8 (10.8) | 160 (35.3) | S[c] |
| DHLA[d] (adapted version) | Liu et al, 2020 [47] | None specified | Taiwan | 1871 | G[e] | NA[f] | 1011 (63.7) | P[g] |
| DHLAT[h] (adapted version) | St. Jean et al, 2017 [53] | None specified | United States | 19 | T (young people aged 12-15 y) | 12.9 (0.9) | 12 (63.2) | P |
| DHLC[i] | Rachmani et al, 2022 [51] | DigComp 2.0 | Indonesia | 383 | G | 37.59 (12.69) | 219 (57.2) | S |
| DHLI[j] | van der Vaart and Drossaert, 2017 [58] | None specified | The Netherlands | 200 | G | 46.4 (19.0) | 107 (53.5) | S and P |
| DHLS[k] | Nelson et al, 2022 [50] | None specified | United States | 508 | T (caregivers of young children) | 34.7 (7.7) | 454 (89.4) | S |
| DHRQ[l] | Scherrenberg et al, 2023 [52] | None specified | Belgium | 315 | T (patients in a clinical context) | 62.6 (15.1) | 118 (37.5) | S |
| DHTL-AQ[m] | Yoon et al, 2022 [33] | Health behavior model and unified theory of acceptance use of technology | South Korea | 590 | T (adults who use digital health technologies) | 46.5 (13.0) | 317 (53.7) | S |
| eHealth Literacy Scale 2.0 | Li, 2018 [44] | Lily Model and social cognitive theory (self-efficacy theory) | United States | 238 | T (low income and socially marginalized population) | 47.1 (13.63) | 175 (75.10) | S |
| eHEALS[n] | Norman and Skinner, 2006 [6] | Lily Model and social cognitive theory (self-efficacy theory) | Canada | 664 | G | 14.95 (1.24) | 294 (44) | S |
| eHEALS+not specified | Neter and Brainin, 2017 [34] | Lily Model | Israel | 82 | T (adults who use the internet for health-related purposes) | 67 (11) | 49 (60) | S and P |
| eHEALS+not specified | van der Vaart et al, 2011 [54] | None specified | The Netherlands | 189 and 88 | G | 52 (11)[o] and 43 (18)[p] | 70 (37)[o] and 43 (49)[p] | S and P |
| eHEALS+not specified | Xie, 2011 [56] | None specified | United States | 124 | T (older adults) | 68.15 (9.00) | 82 (66.1) | S and P |
| eHEALS-E[q] | Petric et al, 2017 [32] | Lily Model and social cognitive theory (self-efficacy theory) | Slovenia | 644 | T (users of online health communities) | 40 | 535 (83) | S |
| eHEALS-6 | Lin et al, 2021 [45] | None specified | China | 215 | T (older adults) | 71.27 (10.82) | 122 (56.7) | S |



XSL•FO
RenderX

| Instrument(s) | Author, year | Theory | Country | Participants included in primary studies, N | Generic instrument or for specific target group | Age (y), mean (SD) or range | Women (%) | Performance based or self-reported |
|---|---|---|---|---|---|---|---|---|
| eHLA[r] | Karnoe et al, 2018 [30] | Lily Model and eHLF[s] | United States | 475 | G | NR[t] | 245 (51.6) | S |
| eHLQ[u] | Kayser et al, 2018 [28] | eHLF | Denmark | 475 | G | NR | —[v] | S |
| eHLS[w] | Hsu et al, 2014 [39] | None specified | Taiwan | 525 | T (college students) | NR | — | S |
| e-HLS[x] | Seçkin et al, 2016 [31] | None specified | United States | 710 | G | 48.82 (16.43) | 381 (53.7) | S |
| eHLS-Web 3.0[y] | Liu et al, 2021 [46] | None specified | China | 421 | T (college students) | 20.5 (1.4) | 218 (51.8) | S |
| eHLT[z]+eHLR[aa] | Camiling, 2019 [36] | Lily Model | Philippines | 274 | T (high school students) | NR | 146 (53.3) | S/P |
| GR-eHEALS[ab] | Marsall et al, 2022 [49] | Lily Model and social cognitive theory (self-efficacy theory) | Germany | 470 | T (German adult population) | 37.16 (13.4) | 332 (70.6) | S |
| HL19-DIGI[ac] | M-POHL, 2021 [48] | None specified | Austria, Belgium, Czech Republic, Denmark, France, Germany, Hungary, Ireland, Israel, Norway, Portugal, Slovakia, and Switzerland | NA | G | NA | — | S |
| PB-mHLS[ad] | Zhang and Li, 2022 [57] | None specified | China | 552 | G | NR | 264 (47.8) | S |
| PRE-HIT[ae] | Koopman et al, 2014 [42] | Lily Model | United States | 200 | T (older patients with chronic diseases) | 54 | 142 (71) | S |
| READHY[af] | Kayser et al, 2019 [35] | None specified | Denmark | 305 | T (patients with chronic diseases) | 58 | 216 ( 70.8) | S |
| RRSA[ag] | Ivanitskaya et al. 2006 [41] | None specified | United States | 308 | T (college students) | 18 - 23 | 237 (77) | S/P |
| RRSA (adapted version) | Ivanitskaya et al, 2010 [40] | None specified | United States | 1914 | T (college students) | NR | 1397 (73) | P |
| RRSA-h[ah] | Hanik and Stellefson, 2011 [38] | Two-process theory of human information processing | United States | 77 | T (undergraduate health education students) | 21.3 (2.0) | 68 (88.3) | P |
| TeHLI[ai] | Paige et al, 2019 [20] | TMeHL[aj] | United States | 283 | T (older people) | 64.34 (10.49) | 160 (56.5) | S |
| Not specified | Chan and Kaufman, 2011 [37] | None specified | United States | 20 | G | 18 - 65 | 14 (70) | P |



XSL•FO
RenderX

| Instrument(s) | Author, year | Theory | Country | Participants included in primary studies, N | Generic instrument or for specific target group | Age (y), mean (SD) or range | Women (%) | Performance based or self-reported |
|---|---|---|---|---|---|---|---|---|
| Not specified [29] | van der Vaart et al, 2013 [29] | None specified | The Netherlands | 15[o] and 16[p] | T (patients with rheumatic diseases) | 56.4 (10.5)[o] and 48.6 (14.2)[p] | 12 (80)[o] and 13 (81)[p] | P |
| Not specified | van Deursen and van Dijk, 2011 [55] | None specified | The Netherlands | 88 | G | 18 - 80 | 43 (49) | P |

[a]CeHLS-D: Condition-specific eHealth literacy Scale for Diabetes.
[b]T: specific target group.
[c]S: self-reported.
[d]DHLA: digital health literacy assessment.
[e]G: generic.
[f]Not available.
[g]P: perfomanced based.
[h]DHLAT: digital health literacy assessment tool.
[i]DHLC: digital health literacy competencies for citizens.
[j]DHLI: digital health literacy instrument.
[k]DHLS: Digital Health Care Literacy Scale.
[l]DHRQ: Digital Health Readiness Questionnaire.
[m]DHTL-AQ: Digital Health Technology Literacy Assessment Questionnaire.
[n]eHEALS: 8-item eHealth Literacy Scale.
[o]Study 1.
[p]Study 2.
[q]eHEALS-E: eHealth Literacy Scale–extended.
[r]eHLA: eHealth literacy assessment toolkit.
[s]eHLF: eHealth literacy framework.
[t]NR: not reported.
[u]eHLQ: eHealth Literacy Questionnaire.
[v]Not applicable.
[w]eHLS: eHealth Literacy Scale.
[x]e-HLS: Electronic Health Literacy Scale.
[y]eHLS-Web 3.0: eHealth Literacy Scale Web 3.0.
[z]eHLT: eHealth Literacy Test.
[aa]eHLR: eHealth Literacy Rubric.
[ab]GR-eHEALS: German eHealth Literacy Scale.
[ac]HL19-DIGI: Health Literacy Survey 19 Digital.
[ad]PB-mHLS: problem-based mHealth Literacy Scale.
[ae]PRE-HIT: The Patient Readiness to Engage in Health Internet Technology instrument.
[af]READHY: Readiness and Enablement Index for Health Technology.
[ag]RRSA: Research Readiness Self-Assessment.
[ah]RRSA-h: Research Readiness Self-Assessment-health scale.
[ai]TeHLI: transactional eHealth literacy instrument.
[aj]TMeHL: transactional model of eHealth literacy.



XSL•FO
RenderX

**Table .** Details of instruments measuring digital health literacy included in the reviews and their frequency of use.

| Review (author, year) | DHLMIs[a] and frequency of their use in primary studies | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | eHEALS[b] [6] (n=163) | eHLQ[c] [28] (n=13) | DHLI[d] [58] (n=12) | eHLA[e] [30] (n=8) | TeHLI[f] [20](n=5) | e-HLS[g] [31](n=3) | DHTL-AQ[h] [33] (n=2) | eHEALS-not specified [34] (n=2) | eHEALS-E [32](n=3) | READ-HY[i] [35] (n=2) | Other [29,36-57] (n=21) | Total |
| Ahn and Kim, 2024 [4] | 3 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | | 7 | 17 |
| Bai et al, 2025 [22] | 57 | 2 | 5 | | | | | | 1 | | 2 | 67 |
| Crocker et al, 2023 [16] | | | | 3 | | | | 1 | | | 10 | 14 |
| Délétroz et al, 2026 [14] | X[1j] | X[1] | X[1] | X[1] | X[1] | X[1] | X[1] | | X[1] | | X[1] | |
| Dijkman et al, 2023 [23] | 12 | 4 | 1 | 1 | | | | | | 1 | | 19 |
| Faux-Nightingale et al, 2022 [11] | 14 | | 1 | 1 | 1 | | | | | | | 17 |
| Huang et al, 2023 [24] | | | | | 1 | | | 1 | | | | 2 |
| Kaihlanen et al, 2023 [25] | 8 | 3 | 1 | 1 | 1 | 1 | 1 | | | 1 | 2 | 19 |
| Lee et al, 2021 [12] | 48 | 2 | 2 | 1 | 1 | 1 | | | 1 | | | 56 |
| Wang and Luan, 2022 [26] | X[1] | X[1] | X[1] | | | X[1] | | | | | | |
| Wang et al, 2025 [15] | X[1] | X[1] | X[1] | X[1] | X[1] | X[1] | X[1] | | X[1] | | X[1] | |
| Xie et al, 2022 [27] | 21 | 1 | 1 | | | | | | | | | 23 |

[a]DHLMI: digital health measurement instrument.

[b]eHEALS: 8-item eHealth literacy scale.

[c]eHLQ: eHealth literacy questionnaire.

[d]DHLI: digital health literacy instrument.

[e]eHLA: eHealth Literacy Assessment toolkit.

[f]TeHLI: Transactional eHealth Literacy Instrument.

[g]e-HLS: Electronic Health Literacy Scale.



XSL•FO
RenderX

[h]DHTL-AQ: Digital Health Technology Literacy Assessment Questionnaire.

[i]READHY: Readiness and Enablement Index for Health Technology.

[j]X[1]: no information on the frequency of the measurement instrument in primary studies.

### Overview of Assigned Dimensions

The number of assigned dimensions per instrument and per original dimension is shown in Figure 2.

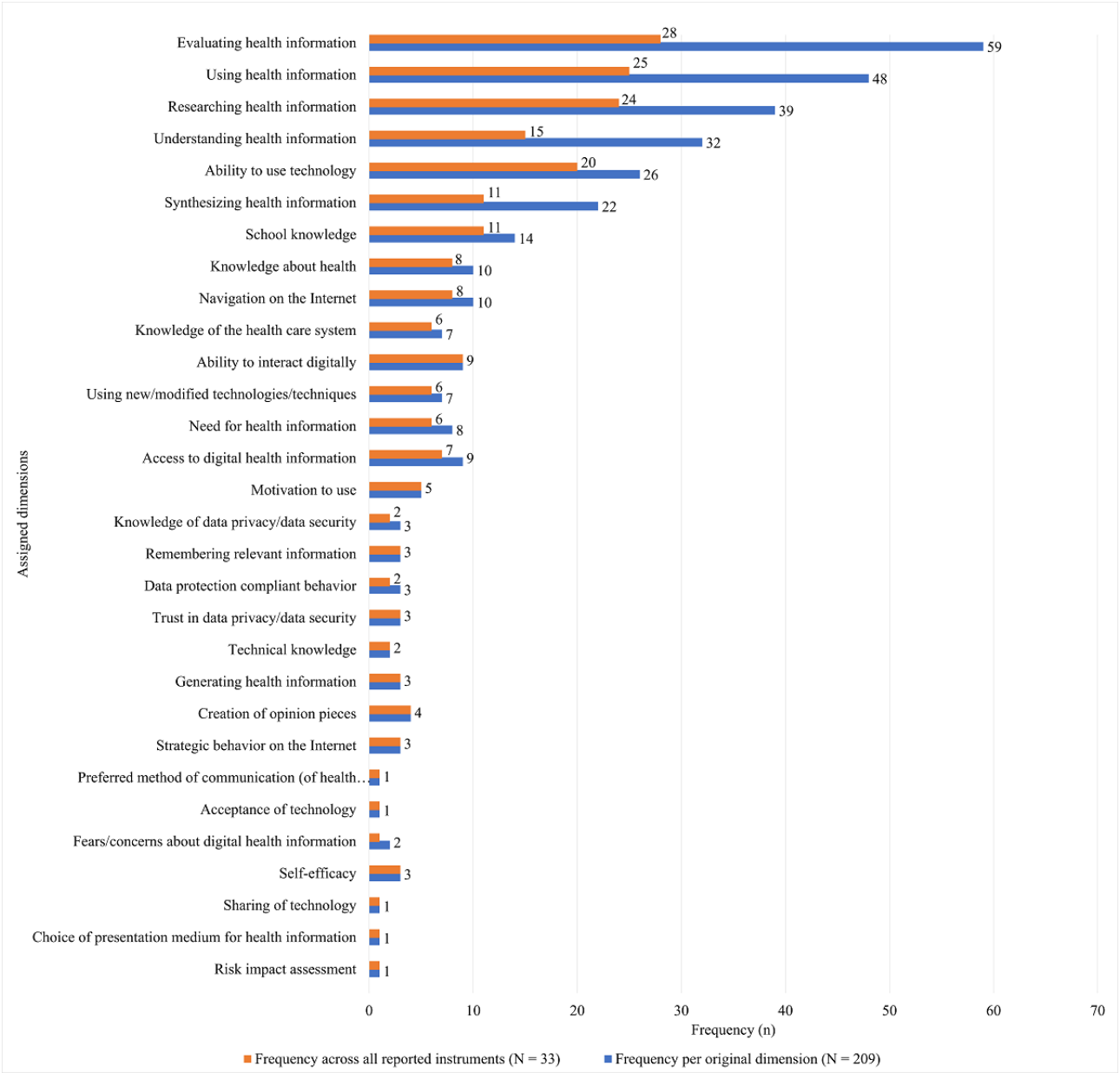
A total of 30 dimensions were assigned during the qualitative content analysis. For some dimensions, a clear categorization was not possible; this occurred in 17 cases for the frequency per original dimension and in 7 cases across all reported instruments. The most frequently assigned dimension was "evaluating health information." This dimension was measured in 28 (84%) of the 33 instruments. This was followed by "using health information" (n=25, 76%), "researching health information" (n=24, 73%), "ability to use technology" (n=20, 61%), and "understanding health information" (n=15, 45%). A very similar order emerged when counting the assigned dimensions according to the original dimensions of the measurement instruments: the dimension "evaluation of health information" was measured most frequently, accounting for 59 of the 209 original dimensions. Consequently, this dimension was operationalized in 59 different ways ("approaches"). The other most commonly reported dimensions were "using health information" (n=48, 23%), "researching health information" (n=39, 19%), "understanding health information" (n=32, 15%), and "ability to use technology" (n=26, 12%).

The assigned dimensions derived from the original dimensions could be sorted into 6 domains (Figure 3): "Action," "Knowledge," "Technology & Privacy," "Preferences/Needs & Motivation," "Risks From Health Information," and "Other." Four of the 5 most used categories (see above) belong to the "Action" domain. The dimensions from the domain "Risks from Health Information" were used least frequently. The domains covered by each instrument are presented in Table 4. A codebook containing the domains, their assigned dimensions, and descriptions is provided in Multimedia Appendix 2.



XSL•FO
RenderX

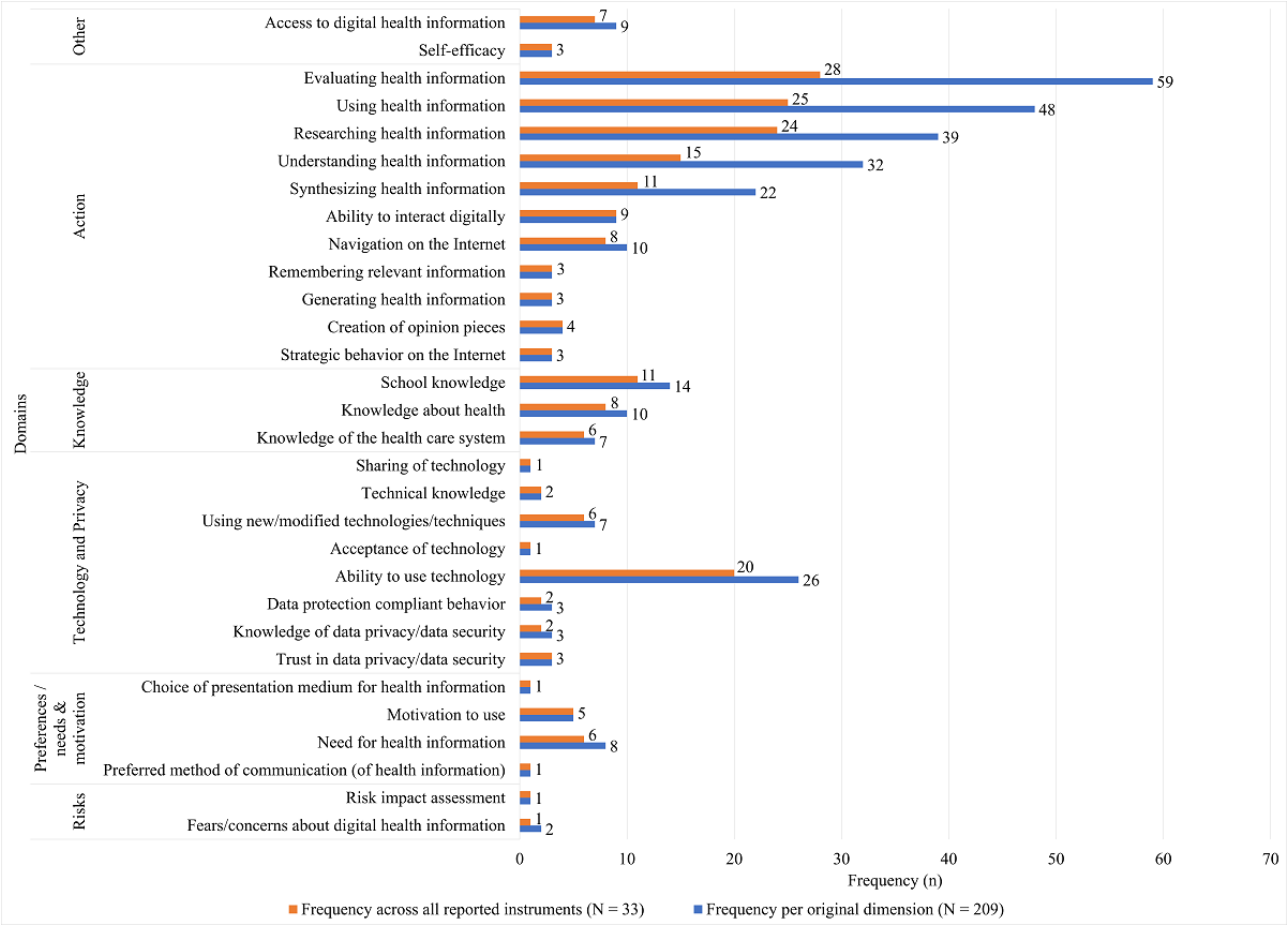
**Figure 2.** Frequency of use of the assigned dimensions.






XSL•FO
RenderX

**Figure 3.** Frequency of use of the assigned dimensions by domains. Action includes skills like understanding, using, and evaluating health information, as well as digital interaction and navigation. Knowledge refers to factual knowledge such as understanding the healthcare system, school-taught content, and general health knowledge. Preferences/needs & motivation encompasses preferred formats, need for health information, and motivation to engage with it. Risks from health information involves concerns and perceived risks related to digital health information. Technology & Privacy covers data protection, trust in digital security, technology use and sharing, and technical knowledge.






XSL•FO
RenderX

**Table .** Categories (domains) of the digital health literacy assessment tools.

| Instruments | Author, year | Action | Technology and privacy | Knowledge | Preferences or needs and motivation | Risks from health information | Categorization not possible | Other |
|---|---|---|---|---|---|---|---|---|
| CeHLS-D[a] | Lee et al, 2022 [43] | X | X | | | | | |
| DHLA[b] (adapted version) | Liu et al, 2020 [47] | X | | X | | | X | |
| DHLAT[c] (adapted version) | St. Jean et al, 2017 [53] | X | | X | | | | |
| DHLC[d] | Rachmani et al, 2022 [51] | X | X | X | | | | |
| DHLI[e] | van der Vaart and Drossaert 2017 [58] | X | X | | | | | |
| DHLS[f] | Nelson et al, 2022 [50] | | | | | | X | |
| DHRQ[g] | Scherrenberg et al, 2023 [52] | X | X | | X | | | |
| DHTL-AQ[h] | Yoon et al, 2022 [33] | X | X | | | | | |
| eHealth Literacy Scale 2.0 | Li, 2018 [44] | X | X | | | | | X |
| eHEALS[i] | Norman and Skinner, 2006 [6] | X | X | X | | | | X |
| eHEALS+not specified | Neter and Brainin, 2017 [34] | X | X | X | | | X | X |
| eHEALS+not specified | van der Vaart et al, 2011 [54] | X | X | X | | | | X |
| eHEALS+not specified | Xie, 2011 [56] | X | X | X | | | | X |
| eHEALS-E[j] | Petric et al, 2017 [32] | X | | | X | | | |
| eHEALS-6 | Lin et al, 2021 [45] | X | | | X | | | X |
| eHLA[k] | Karnoe et al, 2018 [30] | X | X | X | X | | | |
| eHLQ[l] | Kayser et al, 2018 [28] | X | X | X | X | | | |
| eHLS[m] | Hsu et al, 2014 [39] | X | | X | | | | |
| e-HLS[n] | Seçkin et al, 2016[31] | X | | | | | | |
| eHLS-Web 3.0[o] | Liu et al, 2021 [46] | X | | | | | | |
| eHLT[p]+eHLR[q] | Camiling, 2019 [36] | X | X | | | | | |
| GR-eHEALS[r] | Marsall et al, 2022 [49] | X | X | X | X | X | | X |



XSL•FO
RenderX

| Instruments | Author, year | Action | Technology and privacy | Knowledge | Preferences or needs and motivation | Risks from health information | Categorization not possible | Other |
|---|---|---|---|---|---|---|---|---|
| HL19-DIGI[s] | M-POHL, 2021 [48] | X | X | | | | | |
| PB-mHLS[t] | Zhang and Li, 2022 [57] | X | X | X | X | | | X |
| PRE-HIT[u] | Koopman et al, 2014 [42] | | X | | X | X | X | |
| READHY[v] | Kayser et al, 2019 [35] | X | X | X | | | X | X |
| RRSA | Ivanitskaya et al, 2006 [41] | X | X | X | | | X | |
| RRSA[w] (adapted version) | Ivanitskaya et al, 2010 [40] | X | X | X | | | X | |
| RRSA-h[x] | Hanik and Stellefson 2011 [38] | X | | | | | | |
| TeHLI[y] | Paige et al, 2019 [20] | X | X | | | X | | |
| Not specified | Chan and Kaufman, 2011 [37] | X | X | X | X | | | |
| Not specified | van der Vaart et al, 2013 [29] | X | X | | | | | |
| Not specified | van Deursen et al, 2011 [55] | X | X | | | | | |

[a]CeHLS-D: Condition-specific eHealth literacy Scale for Diabetes.
[b]DHLA: digital health literacy assessment.
[c]DHLAT: digital health literacy assessment tool
[d]DHLC: digital health literacy competencies for citizens.
[e]DHLI: digital health literacy instrument.
[f]DHLS: Digital Health Care Literacy Scale.
[g]DHRQ: Digital Health Readiness Questionnaire.
[h]DHTL-AQ: Digital Health Technology Literacy Assessment Questionnaire.
[i]eHEALS: 8-item eHealth Literacy Scale.
[j]eHEALS-E: eHealth Literacy Scale-Extended.
[k]eHLA: eHealth Literacy Assessment Toolkit.
[l]eHLQ: eHealth Literacy Questionnaire.
[m]eHLS: eHealth Literacy Scale.
[n]e-HLS: Electronic Health Literacy Scale.
[o]eHLS-Web 3.0: eHealth Literacy Scale Web 3.0.
[p]eHLT: eHealth Literacy Test.
[q]eHLR: eHealth Literacy Rubric.
[r]GR-eHEALS: German eHealth Literacy Scale.
[s]HL19-DIGI: Health Literacy Survey 19 Digital.
[t]PB-mHLS: Problem-Based mHealth Literacy Scale.
[u]PRE-HIT: The Patient Readiness to Engage in Health Internet Technology Instrument.
[v]READHY: Readiness and Enablement Index for Health Technology.
[w]RRSA: Research Readiness Self-Assessment.
[x]RRSA-h: Research Readiness Self-Assessment-Health Scale.
[y]TeHLI: transactional eHealth literacy instrument.



XSL•FO
RenderX

## Discussion

### Principal Findings

Our analysis identified 33 DHLMIs from 12 reviews. Almost two-thirds (n=20, 61%) of these instruments were not based on a theory. The majority of the instruments (21/33, 64%) were developed for a specific target group. Most of them (20/33, 61%) assessed DHL via self-report, while only 21% (7/33) relied exclusively on performance-based assessment.

The analysis of the measured dimensions revealed a wide variety of DHL dimensions across the instruments. However, the qualitative content analysis revealed that there are key dimensions of DHL that are captured in almost all instruments, but often under different names. These key dimensions include evaluating health information, using health information, researching health information, and the ability to use technology. Such key dimensions covered by a majority of instruments were defined and operationalized differently throughout the included studies and instruments, severely limiting the comparability of results [59]. This problem could be due to a low discriminant validity of health literacy subconstructs [60], of which DHL is one, with navigational health literacy being another, which partly covers similar constructs [61].

The lack of any theory as a basis for several of the measurement instruments analyzed is especially noteworthy because existing models are based on existing evidence and definitions themselves [7]. Thus, the use of acknowledged models for instrument development ensures content validity [62]. Furthermore, comparability of results of different instruments is more feasible when they are based on the same theoretically sound dimensions [63]. That being said, the wide number of assigned dimensions within our content analysis that applied to more than one original dimension, and, likewise, the number of original dimensions that fit into more than one assigned dimension hint at a limited comparability of the results the measurement instruments we studied will likely gain [59].

Comparability of results, however, is the basis for the generalization of study results and distinguishing between target groups is only possible based on a unified set of dimensions on which said groups might differ [64]. It is because of considerations such as this that researchers call for core outcome sets in the measurement of key dimensions to be measured in evaluation studies [65], not only, but also concerning digital health applications.

We acknowledge that DHL, such as health literacy, is a context-dependent, dynamic construct, with the required competencies varying substantially across societies, health care systems, and target populations [66-68]. This contextual dependency, combined with the diversity of target populations (eg, older adults vs students), application contexts (eg, clinical vs community settings), and assessment methods (eg, self-reported vs performance-based measures), makes developing a single, universal DHLMI a significant challenge. Nevertheless, this does not preclude identifying a common conceptual core. A Core Outcome Set of key DHL dimensions would not prescribe a fixed, universal instrument; rather, it would provide a shared dimensional framework that can be adapted to specific cultural contexts, health care systems, and target populations. This approach would enable researchers to use contextually appropriate instruments while ensuring that a common set of dimensions is covered. This would increase the comparability of findings across studies that use different DHLMIs.

Our qualitative content analysis of DHL dimensions builds on and extends the work of Wang et al [15], who also conducted a content analysis of eHL instrument dimensions. However, the 2 analyses differ in scope due to differences in inclusion criteria, which likely explain differences in the resulting dimension landscape.

Wang et al restricted inclusion to studies whose primary focus was the development of a new DHL instrument, explicitly excluding studies in which instrument development was not the main objective. Furthermore, they included only instruments developed for the general population, excluding instruments designed for specific target groups. Our review, in contrast, included both instrument development studies and application studies—for example, studies evaluating digital health interventions or examining DHL as an outcome in clinical or community settings—as well as instruments developed for specific populations, such as older adults or patients with chronic conditions. This broader inclusion allows our analysis to reflect more closely which instruments and dimensions are actually used in research practice, rather than being limited to the subset of instruments and dimensions prominent in the instrument development literature.

As a result, our analysis captures a broader and more heterogeneous range of dimensions, which may partly be explained by the varying constructs and comparison measures used during instrument development and validation. Instruments were often validated against different related concepts, including health literacy, digital literacy, self-efficacy, or technology use, reflecting divergent assumptions regarding the conceptual boundaries of DHL. This may contribute to the considerable heterogeneity observed across existing measurement tools.

Both our analysis and that of Wang et al confirmed the observation that discrepancies in the conceptualization of DHL across instruments limit comparability—a challenge that becomes even more apparent when the full breadth of instruments in active research use is considered. Comparability and synthesis of research findings are essential to provide informed and comprehensive insights into the current state of DHL. Such insights form the foundation for the development of targeted policies and interventions aimed at strengthening DHL at national and international levels. To date, assessments of DHL have relied primarily on separate primary data collections, such as the national survey conducted in Germany [69]. A more efficient and systematic approach would involve aggregating findings from existing studies through meta-analyses, which allow for a structured integration and comparison of results across contexts and populations, or even across regions and countries. Standardizing measurement instruments would also improve comparability and enable earlier identification of trends, gaps in health care, and emerging needs,



XSL•FO
RenderX

which are critical elements for evidence-based policy-making [70].

For the effective implementation of digital health care solutions, it would be both desirable and necessary—although still rarely practiced—for health care providers to assess patients' DHL before suggesting or prescribing the use of such solutions [71]. This consideration motivated our initial intention to provide a short-form measurement instrument suitable for daily health care practice and its time restraints [23]. Systematic assessment would help ensure that digital solutions reach those who can benefit most, thereby improving both individual health outcomes and quality of care while at the same time reducing health disparities due to a digital divide [23].

The large number of measurement instruments available also poses a challenge for developers of digital health solutions, making it difficult to identify and select appropriate tools for assessing the impact of such interventions on DHL [72]. This issue is particularly relevant in the context of effectiveness assessments, such as those required for approval of the German digital health application. The selection of the most appropriate and meaningful instrument to demonstrate effectiveness remains an open question in this process [73]. Furthermore, the development of digital interventions specifically designed to improve, rather than merely assess, DHL requires a solid theory and validated measurement tools that capture the core dimensions of DHL [74].

### Strengths and Limitations

To the best of our knowledge, this is the first review to synthesize both self-reported and performance-based instruments for measuring DHL with a structured qualitative content analysis to systematically analyze and synthesize the dimensions covered by DHLMIs across both instruments developed for the general population and those targeting specific populations. We also provide an overview of available instruments, categorized by target group and a theory. This overview offers a foundation for the refinement of existing or the development of new instruments and serves as practical guidance for researchers, practitioners, and policymakers when selecting suitable tools.

A particular strength of this review is its inclusive conceptual scope. Reflecting the conceptual heterogeneity of DHL, our review systematically mapped all available instruments, not restricting inclusion to those operationalizing health information competencies alone. The resulting dimensional framework therefore extends beyond information-seeking and appraisal to encompass competencies such as technology use, interactive engagement, and the ability to actively generate and share health-related content. These dimensions are directly relevant to the use of digital health services. This broad coverage means that the findings of this review are relevant not only to health information contexts but also to the growing landscape of digital health service use.

However, our review also has limitations. It is neither a full-scale systematic nor an umbrella review [75]. Despite our narrative approach, we have applied several systematic principles in key areas of our work. These include defining a clear research question, applying explicit inclusion and exclusion criteria, describing the selection process for reviews and instruments, conducting structured data extraction, and transparently presenting and discussing the results. Given the systematic nature of our approach in many respects, we considered using the term "review of reviews" but decided against it because our work does not meet all the associated criteria and we did not want to misrepresent our methodology. The chosen narrative format nevertheless seems appropriate, as the research field is conceptually heterogeneous and the included studies vary considerably in methodology and research focus. It is precisely this heterogeneity that precluded a quantitative synthesis of the findings. It should also be noted that the theoretical framework of DHL remains incomplete. Existing theoretical frameworks, including those underlying the 8-item eHealth Literacy Scale and related eHealth competency models, do not yet provide a sufficient basis for the creation of a comprehensive model that covers the many dimensions of this field. When theory is insufficient to capture the complexity of a phenomenon, qualitative and interpretive methods can support concept development. Another limitation is that only measurement instruments that have already been analyzed in a review were taken into account. Therefore, very new measurement instruments, for example, were not included in our analysis. As our study focused on the dimensions of DHL rather than the quality of the measurement instruments, it cannot make any statements about their validity or other quality characteristics. Furthermore, future studies should examine the measuring instruments' test quality criteria to establish a solid basis for further research.

Furthermore, we did not extract data on translated, cross-culturally adapted, or validated versions of the included instruments, as this was not one of the objectives specified in advance for our review. Future reviews of instruments for measuring DHL could systematically document the available language versions and cross-cultural validation studies. This would enable a more comprehensive assessment of the international applicability and dissemination of the instruments and the dimensions of DHL they measure.

Qualitative content analysis was not done independently; however, category formation was done by 2 researchers in open discussion and checked by a third.

When considering the "key dimensions" identified through qualitative content analysis, it is important to acknowledge that they are based on measurement instruments whose validity should be critically examined—particularly those relying solely on self-report [12]. Therefore, while these key dimensions should be interpreted with caution, they still form part of the current evidence base.

Our qualitative approach to analyze the existing instruments allowed us to uncover underlying dimensions of DHL that should be covered by a consolidated instrument in the future. From a purely scientific standpoint, our work underlines the need for instruments that are based on theory as well. Future work on a Core Outcome Set for the evaluation of digital health applications will therefore profit from our work as well.



XSL•FO
RenderX

## Conclusions

Our review has 3 direct implications for research and practice. First, researchers can use the resulting dimension profiles to make more informed and transparent decisions when selecting an instrument suited to their research question and target population. Second, instrument developers gain an evidence-based starting point for the conceptual grounding of new or revised instruments. Third, the identified dimensional core across instruments provides a foundation for the development of a core outcome set for DHL, which could support greater comparability across studies in the future.

Given the contextual dependency of DHL across different societies and health care systems, the goal is not to develop a single universal instrument but to establish a shared conceptual core of key dimensions that can serve as a basis for culturally and population-specific adapted instruments, ultimately supporting greater comparability of findings across diverse research contexts.

## Acknowledgments

We would like to thank Franz Kappert for proofreading the manuscript in terms of understandability and Jesus Escalona for helpful input regarding the qualitative content analysis.

The authors declare the use of generative AI (GenAI) in the research and writing process. According to the Generative AI Delegation Taxonomy [76], the following tasks were delegated to GenAI tools under full human supervision: editing, summarizing text, translation, reformatting, and recommendations. AI tools were not used for any of the following: conceptualization, literature review, methodology, software development and automation, data management, or ethics review. Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

## Funding

The authors would like to thank the Else Kröner Fresenius Center (EKFZ) for Digital Health for funding the research group “Evidence-based Implementation of Digital Health Solutions” in the context of which this review was conducted.

## Authors' Contributions

MS, LH, and JS developed the project proposal together and obtained funding for the project. MS is responsible for the overall project leadership. PK, LH, and MS were responsible for the study design of the review. PK conducted the literature search. PK conducted the data extraction, which was verified for accuracy and completeness by LH and MS. PK and LH performed the qualitative content analysis with input from MS. PK, LH, JS, and MS derived points for discussion and implications. PK drafted the manuscript. All authors critically revised the manuscript for intellectual important content. All authors have read and agreed to the published version of the manuscript.

## Conflicts of Interest

PK declares that she has no conflicting interests. Unrelated to this study, MS received a personal honorarium for speaking engagements from the Saxony Medical Service, the Association of Substitute Health Insurance Funds (Verband der Ersatzkassen e. V.), and the Saxony State Medical Association. Unrelated to this study, LH has received personal grants from Thieme for participation in a discussion board on digital health application in occupational therapy, from the Volkswagenstiftung for participation in a scoping workshop on organizational healthcare research, from the Saxony State Medical Association for a speaking engagement and from Springer Publishing for a first authorship in Das Gesundheitswesen. Unrelated to this study, JS reports institutional grants for investigator-initiated research from the German Federal Joint Committee (G-BA), the Federal Ministry of Health (BMG), the Federal Ministry of Education and Research (BMFTR), the Federal State of Saxony, Novartis, Sanofi, ALK, and Pfizer. He also participated in advisory board meetings as a paid consultant for Sanofi, Lilly, and ALK.

Multimedia Appendix 1
Dimensions of digital health literacy of included measurement instruments.
[DOCX File, 29 KB - i-jmr_v15i1e91603_app1.docx ]

Multimedia Appendix 2
Codebook for domains, assigned dimensions, and descriptions.
[DOCX File, 18 KB - i-jmr_v15i1e91603_app2.docx ]

## References

1. Lee EH, Lee YW, Moon SH. A structural equation model linking health literacy to self-efficacy, self-care activities, and health-related quality of life in patients with type 2 diabetes. Asian Nurs Res (Korean Soc Nurs Sci) 2016 Mar;10(1):82-87. [doi: 10.1016/j.anr.2016.01.005] [Medline: 27021840]



XSL•FO
RenderX

2. Zhang NJ, Terry A, McHorney CA. Impact of health literacy on medication adherence: a systematic review and meta-analysis. Ann Pharmacother 2014 Jun;48(6):741-751. [doi: 10.1177/1060028014526562] [Medline: 24619949]
3. Ernstmann N, Bauer U, Berens EM, et al. DNVF Memorandum Gesundheitskompetenz (Teil 1) – Hintergrund, Relevanz, Gegenstand und Fragestellungen in der Versorgungsforschung. Gesundheitswesen 2020 Jul;82(7):e77-e93. [doi: 10.1055/a-1191-3689]
4. Ahn JW, Kim MY. A systematic review of questionnaire measuring eHealth literacy. Res Community Public Health Nurs 2024;35(3):297. [doi: 10.12799/rcphn.2024.00752]
5. Neter E, Brainin E. Association between health literacy, eHealth literacy, and health outcomes among patients with long-term conditions. Eur Psychol 2019 Jan;24(1):68-81. [doi: 10.1027/1016-9040/a000350]
6. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. J Med Internet Res 2006 Nov 14;8(4):e27. [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
7. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9. [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
8. Ban S, Kim Y, Seomun G. Digital health literacy: a concept analysis. Digit HEALTH 2024;10:20552076241287894. [doi: 10.1177/20552076241287894] [Medline: 39381807]
9. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125. [doi: 10.2196/jmir.2035] [Medline: 22193243]
10. Arias López MDP, Ong BA, Borrat Frigola X, et al. Digital literacy as a new determinant of health: a scoping review. PLOS Digit Health 2023 Oct;2(10):e0000279. [doi: 10.1371/journal.pdig.0000279] [Medline: 37824584]
11. Faux-Nightingale A, Philp F, Chadwick D, Singh B, Pandyan A. Available tools to evaluate digital health literacy and engagement with eHealth resources: a scoping review. Heliyon 2022 Aug;8(8):e10380. [doi: 10.1016/j.heliyon.2022.e10380] [Medline: 36090207]
12. Lee J, Lee EH, Chae D. eHealth literacy instruments: systematic review of measurement properties. J Med Internet Res 2021;23(11):e30644. [doi: 10.2196/30644]
13. Schmitt J, Langan S, Williams HC, European Dermato-Epidemiology Network. What are the best outcome measurements for atopic eczema? A systematic review. J Allergy Clin Immunol 2007 Dec;120(6):1389-1398. [doi: 10.1016/j.jaci.2007.08.011] [Medline: 17910890]
14. Délétroz C, Allen MC, Yameogo AR, et al. Systematic review of the measurement properties of patient-reported outcome measures (PROMs) of eHealth literacy in adult populations. Syst Rev 2026 Jan 24;15(1):66. [doi: 10.1186/s13643-026-03072-6] [Medline: 41580772]
15. Wang C, Chang L, Chen X, Kong J, Qi H. eHealth literacy assessment instruments: scoping review. J Med Internet Res 2025 Aug 20;27:e66965. [doi: 10.2196/66965] [Medline: 40835422]
16. Crocker B, Feng O, Duncan LR. Performance-based measurement of eHealth literacy: systematic scoping review. J Med Internet Res 2023 Jun 2;25:e44602. [doi: 10.2196/44602] [Medline: 37266975]
17. Baumeister RF, Leary MR. Writing narrative literature reviews. Review of General Psychology 1997 Sep;1(3):311-320. [doi: 10.1037/1089-2680.1.3.311]
18. Sukhera J. Narrative reviews: flexible, rigorous, and practical. J Grad Med Educ 2022 Aug;14(4):414-417. [doi: 10.4300/JGME-D-22-00480.1] [Medline: 35991099]
19. Norgaard O, Furstrand D, Klokker L, Karnoe A, Batterham R, Kayser L. The e-health literacy framework: a conceptual framework for characterizing e-health users and their interaction with e-health systems. Knowledge Management & E-Learning 2015 Dec 15;7(4):522-540. [doi: 10.34105/j.kmel.2015.07.035]
20. Paige SR, Stellefson M, Krieger JL, Miller MD, Cheong J, Anderson-Lewis C. Transactional eHealth literacy: developing and testing a multi-dimensional instrument. J Health Commun 2019;24(10):737-748. [doi: 10.1080/10810730.2019.1666940] [Medline: 31583963]
21. Kuckartz U. Qualitative Text Analysis: A Guide to Methods, Practice and Using Software: Sage; 2014.
22. Bai JJ, Mandoh M, Yu YM, Allman-Farinelli M. A scoping review of tools to assess digital health literacy among middle-aged and older adults for application to dietetic practice. Dietetics 2024;3(4):523-554. [doi: 10.3390/dietetics3040037]
23. Dijkman EM, Ter Brake WWM, Drossaert CHC, Doggen CJM. Assessment tools for measuring health literacy and digital health literacy in a hospital setting: a scoping review. Healthcare (Basel) 2023 Dec 20;12(1):11. [doi: 10.3390/healthcare12010011] [Medline: 38200917]
24. Huang YQ, Liu L, Goodarzi Z, Watt JA. Diagnostic accuracy of eHealth literacy measurement tools in older adults: a systematic review. BMC Geriatr 2023 Mar 29;23(1):181. [doi: 10.1186/s12877-023-03899-x] [Medline: 36978033]
25. Kaihlanen AM, Virtanen L, Kainiemi E, Heponiemi T. Professionals evaluating clients' suitability for digital health and social care: scoping review of assessment instruments. J Med Internet Res 2023 Nov 30;25:e51450. [doi: 10.2196/51450] [Medline: 38032707]
26. Wang X, Luan W. Research progress on digital health literacy of older adults: a scoping review. Front Public Health 2022;10:906089. [doi: 10.3389/fpubh.2022.906089] [Medline: 35991040]
27. Xie L, Zhang S, Xin M, Zhu M, Lu W, Mo PKH. Electronic health literacy and health-related outcomes among older adults: a systematic review. Prev Med 2022 Apr;157:106997. [doi: 10.1016/j.ypmed.2022.106997]




XSL•FO
RenderX

28. Kayser L, Karnoe A, Furstrand D, et al. A multidimensional tool based on the eHealth literacy framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36. [doi: 10.2196/jmir.8371] [Medline: 29434011]
29. van der Vaart R, Drossaert CHC, de Heus M, Taal E, van de Laar MAFJ. Measuring actual eHealth literacy among patients with rheumatic diseases: a qualitative analysis of problems encountered using Health 1.0 and Health 2.0 applications. J Med Internet Res 2013 Feb 11;15(2):e27. [doi: 10.2196/jmir.2428] [Medline: 23399720]
30. Karnoe A, Furstrand D, Christensen KB, Norgaard O, Kayser L. Assessing competencies needed to engage with digital health services: development of the eHealth literacy assessment toolkit. J Med Internet Res 2018 May 10;20(5):e178. [doi: 10.2196/jmir.8347] [Medline: 29748163]
31. Seçkin G, Yeatts D, Hughes S, Hudson C, Bell V. Being an informed consumer of health information and assessment of electronic health literacy in a national sample of internet users: validity and reliability of the e-HLS instrument. J Med Internet Res 2016 Jul 11;18(7):e161. [doi: 10.2196/jmir.5496] [Medline: 27400726]
32. Petrič G, Atanasova S, Kamin T. Ill literates or illiterates? Investigating the eHealth literacy of users of online health communities. J Med Internet Res 2017 Oct 4;19(10):e331. [doi: 10.2196/jmir.7372] [Medline: 28978496]
33. Yoon J, Lee M, Ahn JS, et al. Development and validation of digital health technology literacy assessment questionnaire. J Med Syst 2022 Jan 24;46(2):13. [doi: 10.1007/s10916-022-01800-8] [Medline: 35072816]
34. Neter E, Brainin E. Perceived and performed eHealth literacy: survey and simulated performance test. JMIR Hum Factors 2017 Jan 17;4(1):e2. [doi: 10.2196/humanfactors.6523] [Medline: 28096068]
35. Kayser L, Rossen S, Karnoe A, et al. Development of the multidimensional Readiness and Enablement Index for Health Technology (READHY) tool to measure individuals' health technology readiness: initial testing in a cancer rehabilitation setting. J Med Internet Res 2019 Feb 12;21(2):e10377. [doi: 10.2196/10377] [Medline: 30747717]
36. Camiling MKS. eHealth literacy of high school students in the Philippines. ije 2019;7(2):69-87. [doi: 10.22492/ije.7.2.04]
37. Chan CV, Kaufman DR. A framework for characterizing eHealth literacy demands and barriers. J Med Internet Res 2011 Nov 17;13(4):e94. [doi: 10.2196/jmir.1750] [Medline: 22094891]
38. Hanik B, Stellefson M. E-Health literacy competencies among undergraduate health education students: a preliminary study. Int Electron J Health Educ 2011;14:46-58 [FREE Full text]
39. Hsu W, Chiang C, Yang S. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014 Dec 12;16(12):e287. [doi: 10.2196/jmir.3542] [Medline: 25499086]
40. Ivanitskaya L, Brookins-Fisher J, O Boyle I, Vibbert D, Erofeev D, Fulton L. Dirt cheap and without prescription: how susceptible are young US consumers to purchasing drugs from rogue internet pharmacies? J Med Internet Res 2010 Apr 26;12(2):e11. [doi: 10.2196/jmir.1520] [Medline: 20439253]
41. Ivanitskaya L, O'Boyle I, Casey AM. Health information literacy and competencies of information age students: results from the interactive online Research Readiness Self-Assessment (RRSA). J Med Internet Res 2006 Apr 21;8(2):e6. [doi: 10.2196/jmir.8.2.e6] [Medline: 16867969]
42. Koopman RJ, Petroski GF, Canfield SM, Stuppy JA, Mehr DR. Development of the PRE-HIT instrument: patient readiness to engage in health information technology. BMC Fam Pract 2014 Jan 28;15(18):18. [doi: 10.1186/1471-2296-15-18] [Medline: 24472182]
43. Lee EH, Lee YW, Lee KW, et al. Development and psychometric evaluation of a new brief scale to measure eHealth literacy in people with type 2 diabetes. BMC Nurs 2022 Nov 4;21(1):297. [doi: 10.1186/s12912-022-01062-2] [Medline: 36333750]
44. Li X. Understanding eHealth literacy from a privacy perspective: eHealth literacy and digital privacy skills in American disadvantaged communities. American Behavioral Scientist 2018 Sep;62(10):1431-1449. [doi: 10.1177/0002764218787019]
45. Lin Z, Zhang Y, Matteson M, et al. Older adults' eHealth literacy and the role libraries can play. Journal of Librarianship and Information Science 2021 Sep;53(3):488-498. [doi: 10.1177/0961000620962847]
46. Liu HX, Chow BC, Liang W, Hassel H, Huang YW. Measuring a broad spectrum of eHealth skills in the Web 3.0 context using an eHealth literacy scale: development and validation study. J Med Internet Res 2021 Sep 23;23(9):e31627. [doi: 10.2196/31627] [Medline: 34554098]
47. Liu P, Yeh LL, Wang JY, Lee ST. Relationship between levels of digital health literacy based on the Taiwan digital health literacy assessment and accurate assessment of online health information: cross-sectional questionnaire study. J Med Internet Res 2020 Dec 21;22(12):e19767. [doi: 10.2196/19767] [Medline: 33106226]
48. M-POHL. International report on the methodology, results, and recommendations of the european health literacy population survey 2019-2021 (HLS19) of m-POHL. : Austrian National Public Health Institute; 2021 URL: https://m-pohl.net/sites/m-pohl.net/files/inline-files/HLS19%20International%20Report.pdf [accessed 2026-07-29]
49. Marsall M, Engelmann G, Skoda EM, Teufel M, Bäuerle A. Measuring electronic health literacy: development, validation, and test of measurement invariance of a revised German version of the eHealth literacy scale. J Med Internet Res 2022 Feb 2;24(2):e28252. [doi: 10.2196/28252] [Medline: 35107437]
50. Nelson LA, Pennings JS, Sommer EC, Popescu F, Barkin SL. A 3-item measure of digital health care literacy: development and validation study. JMIR Form Res 2022 Apr 29;6(4):e36043. [doi: 10.2196/36043] [Medline: 35486413]




XSL•FO
RenderX

51. Rachmani E, Haikal H, Rimawati E. Development and validation of digital health literacy competencies for citizens (DHLC), an instrument for measuring digital health literacy in the community. Comput Methods Programs Biomed Update 2022;2:100082. [doi: 10.1016/j.cmpbup.2022.100082] [Medline: 36407680]
52. Scherrenberg M, Falter M, Kaihara T, et al. Development and internal validation of the Digital Health Readiness Questionnaire: prospective single-center survey study. J Med Internet Res 2023 Mar 10;25:e41615. [doi: 10.2196/41615] [Medline: 36897627]
53. St. Jean B, Greene Taylor N, Kodama C, Subramaniam M. Assessing the digital health literacy skills of tween participants in a school-library-based after-school program. J Consum Health Internet 2017 Jan 2;21(1):40-61. [doi: 10.1080/15398285.2017.1279894]
54. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 9;13(4):e86. [doi: 10.2196/jmir.1840] [Medline: 22071338]
55. van Deursen AJAM, van Dijk JAGM. Internet skills performance tests: are people ready for eHealth? J Med Internet Res 2011 Apr 29;13(2):e35. [doi: 10.2196/jmir.1581] [Medline: 21531690]
56. Xie B. Experimenting on the impact of learning methods and information presentation channels on older adults' e-health literacy. J Am Soc Inf Sci 2011 Sep;62(9):1797-1807. [doi: 10.1002/asi.21575]
57. Zhang L, Li P. Problem-based mHealth Literacy Scale (PB-mHLS): development and validation. JMIR Mhealth Uhealth 2022 Apr 8;10(4):e31459. [doi: 10.2196/31459] [Medline: 35394446]
58. van der Vaart R, Drossaert C. Development of the digital health literacy instrument: measuring a broad spectrum of Health 1.0 and Health 2.0 Skills. J Med Internet Res 2017 Jan 24;19(1):e27. [doi: 10.2196/jmir.6709] [Medline: 28119275]
59. Brady KJS, Trockel MT, Khan CT, et al. What do we mean by physician wellness? A systematic review of its definition and measurement. Acad Psychiatry 2018 Feb;42(1):94-108. [doi: 10.1007/s40596-017-0781-6] [Medline: 28913621]
60. Rouquette A, Rigal L, Mancini J, et al. Health literacy throughout adolescence: invariance and validity study of three measurement scales in the general population. Patient Educ Couns 2022 Apr;105(4):996-1003. [doi: 10.1016/j.pec.2021.07.044] [Medline: 34384639]
61. Griese L, Berens EM, Nowak P, Pelikan JM, Schaeffer D. Challenges in navigating the health care system: development of an instrument measuring navigation health literacy. Int J Environ Res Public Health 2020 Aug 8;17(16):5731. [doi: 10.3390/ijerph17165731] [Medline: 32784395]
62. Sireci SG. The construct of content validity. Soc Indic Res 1998 Nov;45(1-3):83-117. [doi: 10.1023/A:1006985528729]
63. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117. [doi: 10.2196/13117] [Medline: 31115340]
64. Graham JM, Diebels KJ, Barnow ZB. The reliability of relationship satisfaction: a reliability generalization meta-analysis. J Fam Psychol 2011 Feb;25(1):39-48. [doi: 10.1037/a0022441] [Medline: 21355645]
65. Mathioudakis AG, Khaleva E, Fally M, et al. Core outcome sets, developed collaboratively with patients, can improve the relevance and comparability of clinical trials. Eur Respir J 2023 Apr;61(4):2202107. [doi: 10.1183/13993003.02107-2022]
66. Ishikawa H, Miyawaki R, Kato M, Muilenburg JL, Tomar YA, Kawamura Y. Digital health literacy and trust in health information sources: a comparative study of university students in Japan, the United States, and India. SSM Popul Health 2025 Sep;31:101844. [doi: 10.1016/j.ssmph.2025.101844] [Medline: 40746760]
67. Shaw SJ, Huebner C, Armin J, Orzech K, Vivian J. The role of culture in health literacy and chronic disease screening and management. J Immigr Minor Health 2009 Dec;11(6):460-467. [doi: 10.1007/s10903-008-9135-5] [Medline: 18379877]
68. Shi Z, Du X, Li J, Hou R, Sun J, Marohabutr T. Factors influencing digital health literacy among older adults: a scoping review. Front Public Health 2024;12:1447747. [doi: 10.3389/fpubh.2024.1447747] [Medline: 39555039]
69. Schaeffer D, Gille S, Berens EM, et al. Digital health literacy of the population in Germany: results of the HLS-GER 2. Gesundheitswesen 2023 Apr;85(4):323-331. [doi: 10.1055/a-1670-7636] [Medline: 34905785]
70. Scheithauer S, Hoffmann J, Lang C, et al. Pandemic preparedness - a proposal for a research infrastructure and its functionalities for a resilient health research system. Gesundheitswesen 2025 Dec;87(S 03):S334-S343. [doi: 10.1055/a-2365-9179] [Medline: 39009032]
71. Collins SA, Currie LM, Bakken S, Vawdrey DK, Stone PW. Health literacy screening instruments for eHealth applications: a systematic review. J Biomed Inform 2012 Jun;45(3):598-607. [doi: 10.1016/j.jbi.2012.04.001] [Medline: 22521719]
72. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30. [doi: 10.2196/jmir.4055] [Medline: 25639757]
73. König IR, Mittermaier M, Sina C, et al. Evidence of positive care effects by digital health apps-methodological challenges and approaches. Inn Med (Heidelb) 2022 Dec;63(12):1298-1306. [doi: 10.1007/s00108-022-01429-2] [Medline: 36279007]
74. Harst L, Wollschlaeger B, Birnstein J, Fuchs T, Timpel P. Evaluation is key: providing appropriate evaluation measures for participatory and user-centred design processes of healthcare IT. Int J Integr Care 2021 Jun 21;21(2):24. [doi: 10.5334/ijic.5529] [Medline: 34220388]




XSL•FO
RenderX

75. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc 2015 Sep;13(3):132-140. [doi: 10.1097/XEB.0000000000000055] [Medline: 26360830]
76. Suchikova Y, Tsybuliak N, Teixeira da Silva JA, Nazarovets S. GAIDeT (Generative AI Delegation Taxonomy): a taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. Account Res 2026 Apr;33(3):2544331. [doi: 10.1080/08989621.2025.2544331] [Medline: 40781729]

## Abbreviations

**DHL:** digital health literacy
**DHLMI:** digital health measurement instrument
**DiGA:** digital health applications
**TMeHL:** Transactional Model of eHealth Literacy

*Edited by M Balcarras; submitted 23.Jan.2026; peer-reviewed by H Yin, M Hoffrén-Mikkola; revised version received 26.May.2026; accepted 10.Jun.2026; published 28.Aug.2026.*

*Please cite as:*
*Kaboth P, Harst L, Schmitt J, Scheibe M*
*How Many Instruments Do We Need? Strengths, Weaknesses, and Potentials for Further Developments of Digital Health Literacy Measurement Instruments: Narrative Overview of Reviews and Content Analysis*
*Interact J Med Res 2026;15:e91603*
*URL: https://www.i-jmr.org/2026/1/e91603*
*doi:10.2196/91603*

© Pauline Kaboth, Lorenz Harst, Jochen Schmitt, Madlen Scheibe. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 28.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Landscape, Evidence, Gaps, and Opportunities in Digital Mental Health Interventions for Older Adults: Scoping Review

Leyi Zhou[1], MSW; Xinyi Zuo[2], MEd, EdD; Joonyoung Cho[3], MSW, MPH, PhD; Chuxuan Zheng[4], BA; Shengyi Jing[2], MSW; Xiaoling Xiang[2], MSW, PhD

[1]School of Social Welfare, University of California, Berkeley, Berkeley, CA, United States

[2]School of Social Work, University of Michigan, 1080 South University Avenue, Ann Arbor, MI, United States

[3]Center on Aging, Thompson School of Social Work and Public Health, University of Hawaii System, Honolulu, HI, United States

[4]Department of Psychology, College of Literature, Science, and the Arts, University of Michigan, Ann Arbor, MI, United States

**Corresponding Author:**
Xiaoling Xiang, MSW, PhD
School of Social Work, University of Michigan, 1080 South University Avenue, Ann Arbor, MI, United States

## Abstract

**Background:** Mental health conditions, including depression, anxiety, and psychological distress, are prevalent among the aging population and affect their health, functioning, and quality of life. Access to proper and high-quality mental health treatment is necessary; however, mental health treatment and care remain underused due to stigma, workforce shortages, cost, and mobility limitations. Digital mental health interventions (DMHIs) are emerging as a promising strategy to improve the accessibility and effectiveness of mental health services for older adults, but older adults have historically been underrepresented in DMHI development and evaluation. Additionally, the effectiveness of different types of DMHIs and how age-centered design approaches influence outcomes remain underexplored.

**Objective:** This scoping review mapped and synthesized evidence on DMHIs focused on adults aged 50 years and older and identified gaps in the evidence base related to study design, age-related adaptations, and clinical outcomes. Specifically, we examined (1) the technologies and therapeutic approaches used, (2) the outcomes and effectiveness of DMHIs, and (3) age-centered adaptations and their outcomes.

**Methods:** This scoping review searched for studies focusing on DMHIs for older adults across PubMed, PsycINFO, Scopus, Ageline, and Web of Science that were published from 2000 to February 2025. Eligible studies evaluated or described the design of DMHIs targeting mental health conditions among adults aged 50 years or older. Two rounds of independent screening and data extraction were conducted by multiple reviewers. Extracted data included study design, sample characteristics, intervention features, technologies used, age-related adaptations, and clinical outcomes.

**Results:** Seventy-two studies met the inclusion criteria, of which 36 were randomized controlled trials and 54 reported clinical outcomes. Web-based cognitive behavioral therapy was the most commonly used approach, followed by games, virtual reality, mobile apps, chatbots, and robots. Fifty-four studies reported positive clinical outcomes, most commonly reductions in depression, anxiety, or psychological distress. However, only one-third of the studies incorporated age-centered design adaptations or co-design approaches, such as simplified interfaces, larger fonts, age-relevant content, or participatory development with older adults.

**Conclusions:** Among studies reporting positive clinical outcomes, DMHIs can reduce depression, anxiety, and psychological distress. However, with only half of the included studies using randomized controlled trial designs, the overall evidence base remains moderate. In addition, age-adaptive design remains underdeveloped. Future research should strengthen trial designs and systematically examine how usability and age-centered adaptations influence DMHI effectiveness.

***(Interact J Med Res 2026;15:e92542)*** doi:10.2196/92542

**KEYWORDS**

digital mental health interventions; gerontechnology; scoping review; age-adaptation design; cognitive behavioral therapy

## Introduction

Mental health conditions, such as depression and anxiety, are prevalent among older adults and can significantly impact their health and quality of life [1]. Although many evidence-based mental health treatments exist, they are persistently underused among older adults [1]. Factors such as stigma [2], transportation barriers [3], a shortage of mental health professionals [4], financial constraints [5], and insufficient insurance coverage [6] are common access barriers faced by many vulnerable populations needing mental health treatments. These access barriers are compounded in older adults due to misconceptions



XSL•FO
RenderX

about mental health conditions, such as the belief that depression is a normal part of aging [7], challenges in identifying and diagnosing certain medical health conditions in later life [1], and more pervasive mental health stigma [8], as well as reluctance to seek help [9].

Digital mental health interventions (DMHIs) are a promising solution to address the growing unmet mental health needs among older adults. Various technologies have been used to deliver mental health treatments, such as smartphone apps, telecommunication tools (eg, SMS text messaging, emailing, and video conferencing), online platforms, computer-based programs, remote monitoring devices, wearable technologies, and virtual reality (VR) [10-13]. DMHIs have been hailed for their scalability, accessibility, efficiency, and potential cost-effectiveness [10,14]. Many DMHIs can be used repeatedly with different patients without losing their therapeutic power and at a reduced marginal cost, making them especially helpful in underresourced settings [15]. Many DMHIs are self-guided, which may provide more privacy for those with internalized stigma toward mental illness and make them more receptive to treatment [16]. In addition, many DMHIs can be accessed anywhere with an internet connection, thus eliminating the need for attending office-based treatments, a significant barrier for those with transportation barriers due to mobility difficulties or driving cessation [17]. Some DMHI modalities, such as internet-based cognitive behavioral therapy (CBT), have a robust body of evidence supporting their effectiveness in mixed-age samples and diverse settings [18-20].

Although the potential of DMHIs is increasingly recognized, older adults have traditionally been underrepresented and sometimes excluded from DMHI development and testing. Despite the age-related digital divide, older adults have increasingly used technologies for health promotion and maintenance, particularly since the COVID-19 pandemic [21]. The increased adoption and acceptance of technology have paved the way for a wider uptake of DMHIs among older adults. Indeed, we have observed an uptick in DMHI studies focused on older adults, including those from socioeconomically disadvantaged backgrounds [22-24]. Given the fast pace of technological advancement and continued interest in DMHIs, a comprehensive and up-to-date review is essential to pinpoint existing gaps and suggest future directions.

The purpose of this scoping review was to systematically scope and map existing evidence on DMHIs focused on older adults. Previous review papers have focused on DMHI trials to evaluate whether they are clinically effective in reducing depression, anxiety, or other mental health symptoms among older adults [25,26]. Different from reviews that primarily provide evidence on intervention effectiveness, this review emphasizes that the success of DMHIs for older adults depends not only on clinical efficacy but also on usability, age-relevant content, and fit with older adults' needs. This review provides more insight into how DMHIs are designed, how age-centered adaptations are incorporated, and how these adaptations may contribute to positive clinical outcomes. Our goal is to move beyond simply asking whether DMHIs benefit older adults. Instead, we aim to stimulate discussions around which program features are most likely to induce behavioral change, identify opportunities for treatment innovation, and provide insights for designing future DMHIs that better meet the needs of older adults.

To this end, this scoping review aimed to (1) describe the types of technologies and program features used in DMHIs for treating and preventing mental health conditions in older adults, (2) assess the range and nature of procedural, content, and technological adaptations made in DMHIs specifically for older adults, and (3) qualitatively explore and summarize the clinical outcomes of DMHIs across various technologies and program features.

## Methods

### Study Design

A scoping review fits our study aim well, as it allows for a comprehensive mapping of existing studies, identification of gaps in the current literature, summarization of research activities, and dissemination of findings to inform future research, policy, and practice [27,28]. We carried out the review process following the framework developed by Arksey and O'Malley [27] and later refined by Levac et al [28].

### Search Strategy

Electronic searches for this study were initially conducted in June 2022 across 5 databases: PubMed, PsycINFO, Scopus, Ageline, and Web of Science. An updated literature search was performed in February 2025 to capture new studies published since the initial search. Multimedia Appendix 1 shows the search terms, which were developed based on the PICO (population, intervention, comparator, and outcome) framework: (1) population: studies involving participants aged 50 years or older; (2) intervention: DMHIs providing treatment and support for mental health conditions; (3) comparator: any comparator or no comparator; (4) outcome: types of technologies used, adaptations made for older adults, and the effects of DMHIs.

Our search was limited to publication years from January 2000 to February 2025, as studies on telepsychiatry began to emerge around 2000 [29]. Only studies reported in English were included. The references of the included studies and relevant review papers were checked for additional relevant studies. If full-text studies were not available, the corresponding authors were contacted. The initial search process was carried out by the first author, and duplicates were removed prior to screening. The inclusion and exclusion criteria, as well as data extraction forms, were drafted by the first author and the corresponding author.

### Inclusion and Exclusion Criteria

Table 1 presents the inclusion and exclusion criteria for study selection. Studies were included if they: (1) were empirical studies (qualitative or quantitative) evaluating the outcomes of DMHIs or describing the design process of DMHIs; (2) focused on older adults, defined as studies in which all participants were aged 50 years or older, or where analyses were stratified to isolate treatment effects for this age group; (3) involved interventions targeting mental health conditions commonly experienced by older adults; (4) focused on DMHIs, including web-based interventions, software applications, artificial



XSL•FO
RenderX

intelligence (AI) tools, video game consoles, or wearable technology aimed at addressing mental health conditions; and (5) were written in English.



XSL•FO
RenderX

**Table .** Inclusion and exclusion criteria.

| Element | Inclusion criteria | Exclusion criteria |
|---|---|---|
| Population | • At least any kind of mental health conditions (eg, depression, general psychological distress, psychological distress with physical condition, anxiety, posttraumatic stress disorder, and grief)<br>• Aged ≥50 years | • Aged <50 years |
| Intervention | • DMHI[a] (eg, web-based interventions, software apps, artificial intelligence, video game consoles, and wearable technology) | • N/A[b] |
| Comparison | • Different interventions (eg, delayed-treatment waitlist control group, waitlist control group, standardized institutional rehabilitation, support from voluntary caregiver associations, placebo control exercise, and conventional care) | • N/A |
| Outcomes | • Types of technologies being applied, adaptations to older adults, and effects of DMHIs | • N/A |
| Study design | • Empirical studies (qualitative or quantitative) (eg, RCT[c], single-group studies, observational studies, quasi-experimental studies, and mixed methods studies) | • Review articles<br>• Protocols<br>• Perspective pieces |
| Language | • English papers | • Non-English papers |
| Types of mental health problems | • Depression<br>• General psychological distress<br>• Psychological distress with physical condition<br>• Anxiety<br>• Posttraumatic stress disorder<br>• Grief | • Loneliness<br>• Isolation<br>• Sleep disorder |
| Type of technology | • DMHIs (eg, web-based interventions, software apps, artificial intelligence, video game consoles, and wearable technology for any mental health condition) | • Technologies for monitoring mental health symptoms<br>• Focused on cognitive training or cognitive rehabilitation with the primary goal of improving cognition or managing symptoms of dementia<br>• Chronic disease self-management tools with the primary purpose of improving self-management behaviors among persons with chronic physical health conditions<br>• Primarily aimed to improve health and physical functioning, even if they reported improved mental health outcomes<br>• Telepsychiatry or tele–mental health studies, where services were provided by phone or teleconferencing tools by a human therapist<br>• Focused on electronic health or medical records, decision support tools for clinicians, analytic services, services that primarily provided support and education to health professionals, clinical practice management software, and clinical workflow and communication software |

[a]DMHI: digital mental health interventions.

[b]N/A: not applicable.

[c]RCT: randomized controlled trial.



XSL•FO
RenderX

Studies were excluded if they met any of the following criteria: (1) the study sample included individuals under the age of 50 years, and the treatment effects for older adults (≥50 y) were not isolated through stratified analysis; (2) they were review articles, study protocols, or perspective or commentary pieces; (3) the focus was solely on technologies used for monitoring mental health symptoms without a therapeutic component; (4) the primary aim was cognitive training or rehabilitation focused on improving cognition or managing symptoms of dementia; (5) the intervention centered on chronic disease self-management tools with the primary goal of improving self-management behaviors or physical health outcomes, even if mental health outcomes were reported as secondary outcomes; (6) the study focused exclusively on telepsychiatry or tele–mental health interventions delivered by a human therapist through phone or video conferencing; or (7) the study focused on tools such as electronic health or medical records, decision support tools for clinicians, analytic services, educational or support services for health professionals, clinical practice management software, or clinical workflow and communication tools not intended for direct mental health interventions.

In the first round of screening conducted in 2022, 2 reviewers (LZ and JC) independently screened titles and abstracts using the inclusion and exclusion criteria outlined above. A second round was carried out by 2 coauthors (LZ and SJ) in 2025, following the same procedure. Studies meeting all inclusion criteria were retained, and those meeting any exclusion criteria were removed. Disagreements between reviewers were resolved after consultation with the corresponding author.

### Data Extraction and Synthesis

Three reviewers (LZ, JC, and CZ) independently extracted data, with discrepancies resolved by the corresponding author. Data were entered and managed in an electronic database. Extracted data included the country of study, recruitment settings, study design, sample characteristics (eg, sample size, mean age, gender composition, and education level), mental health condition of focus, types of technologies, intervention features (eg, mode of delivery, type of device required, theoretical underpinnings, and treatment duration), clinical outcomes, and a narrative summary of key findings.

We summarized study characteristics and created cross-tabulations of key characteristics to conduct a qualitative synthesis and identify common themes. Specifically, we examined the distinctive features of different types of DMHIs, the mental health conditions they targeted, and the extent to which these technologies improved participants' experiences and outcomes.

## Results

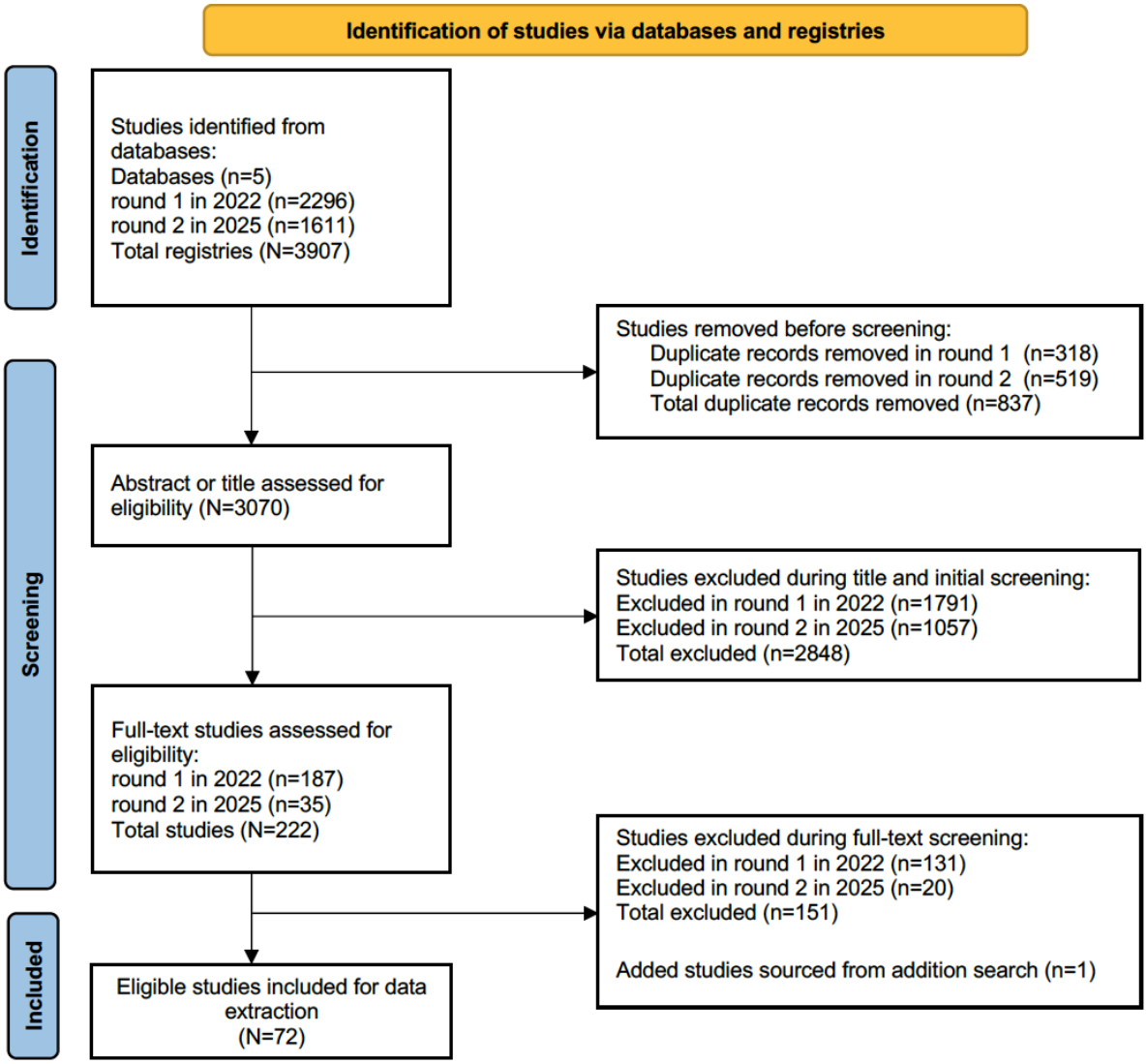
### Study Selection

Figure 1 presents the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the screening process. This scoping review was conducted and reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines (Checklist 1). A total of 3907 potentially relevant studies were identified across rounds 1 and 2, with 3070 remaining after duplicates were removed. Title and abstract screening yielded 222 studies for full-text review. After the full-text review, 72 studies were included in this scoping review. The screening process was managed using the Rayyan platform (Rayyan Systems Inc).



XSL•FO
RenderX

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.




Note on detailed reasons for study exclusion during full-text screening:

- Included participants aged <50 years (Round 1: 73; Round 2: 4; total: n=77)
- Intervention not focused on mental health (eg, focused on physical health, cognitive training, sleep, loneliness or social isolation, or self-management; round 1: 38; round 2: 9; total: n=47)
- Wrong study types (eg, protocol, review, perspective, correspondence, or conference poster; round 1: 8; round 2: 1; total: n=9)
- Not an empirical intervention study (eg, theoretical or conceptual therapeutic paper; round 1: 4; round 2: 0; total: n=4)
- Not a digital mental health intervention (eg, telehealth or SMS text messaging provided by a human therapist; round 1: 5; round 2: 4; total: n=9)
- Study not found (Round 1: 1; Round 2: 2; total: n=3)

## Main Characteristics of the Included Studies

Among the 72 included studies, the majority were published after 2015. Table 2 and Figure 2 illustrate an increasing trend in publications since 2020, reflecting the significant expansion and adoption of digital mental health solutions since the COVID-19 pandemic [30].

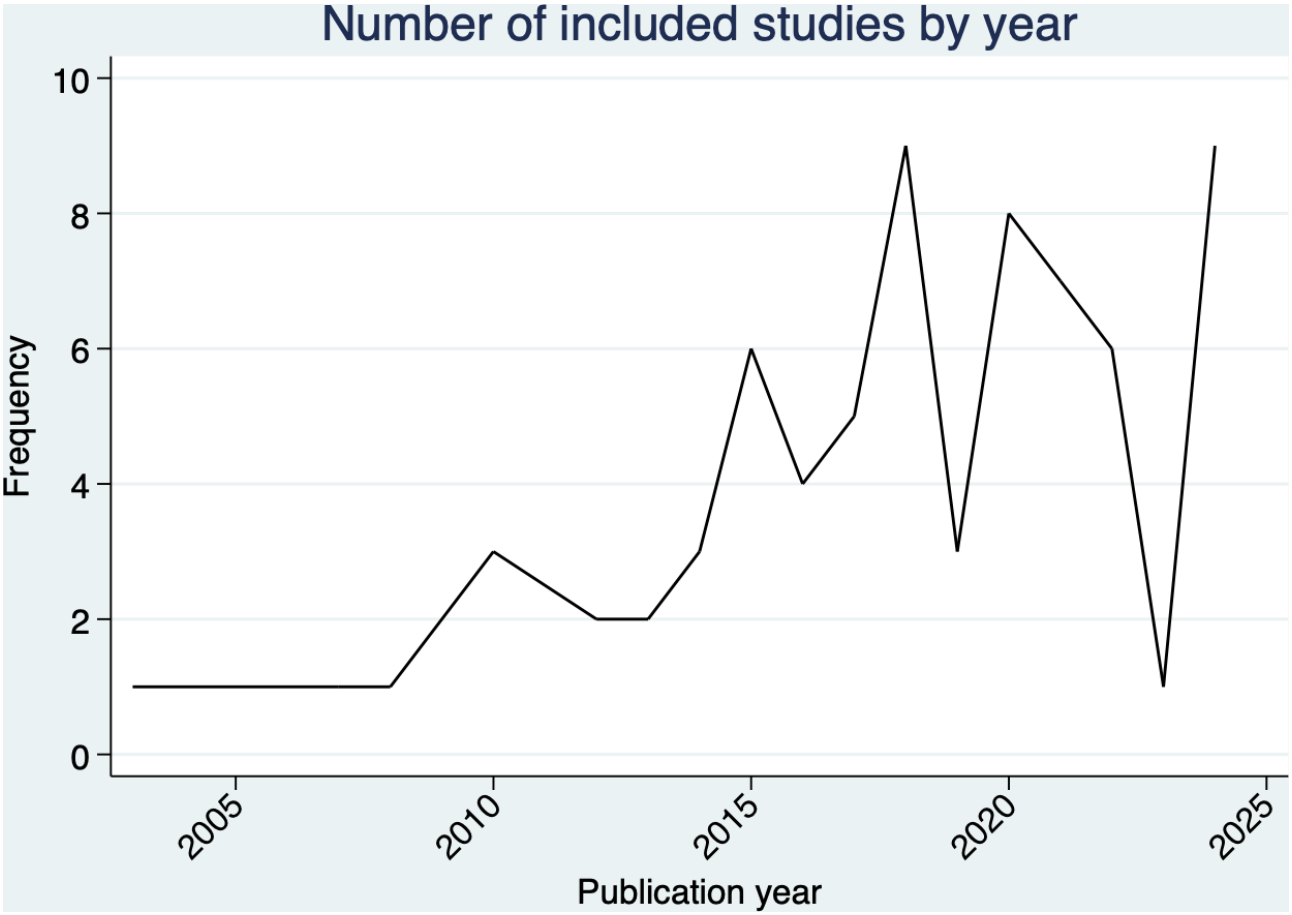


XSL•FO
RenderX

**Table .** Number of included studies per year (N=72).

| Year group | Value, n (%) |
|---|---|
| Before and in 2004 | 1 (1.39) |
| 2005 - 2009 | 3 (4.17) |
| 2010 - 2014 | 10 (13.89) |
| 2015 - 2019 | 27 (37.50) |
| 2020 - 2025 | 31 (43.06) |

**Figure 2.** Number of included studies published each year.




More detailed study characteristics are presented in Multimedia Appendix 2. The intervention took place across 18 countries, with the largest number of studies conducted in the United States (n=31), followed by Australia (n=10). Other top countries included China, Sweden, Italy, and Finland. This indicates that while the United States and Australia have a strong presence, interest in DMHIs is growing globally. That said, most of these countries represent the Global North. The sample sizes of the included studies ranged from as few as 3 participants to over 900. Most studies targeted older adults aged 60 years and above, with mean sample ages ranging from 55 to 88.2 years. Of the 72 studies that reported gender composition, 65 included more than 50% female participants. Most studies recruited participants through a variety of channels, including long-term care agencies, clinical settings, community centers, and online platforms or websites. Completion rates were generally high, with 47 studies reporting adherence rates of 80% or higher.

Regarding study design, nearly half of the studies (36/72) used randomized controlled trials (RCTs), followed by 14 single-group studies, 10 observational studies, 9 quasi-experimental studies, 2 case studies, and 1 mixed methods study. Among the 36 RCTs, 12 used a waitlist control (WLC), 10 used treatment as usual (TAU), 8 used an active comparator (AC), 2 used delayed treatment, 2 used a placebo control, and 2 combined WLC with AC. Of the 9 quasi-experimental studies, 2 used WLC, 2 used TAU, 4 used AC, and 1 used a combination of TAU and AC. RCTs are generally considered the strongest design for establishing causal effects [31]. Within this group, studies using an AC provide the most rigorous evidence, as they control for nonspecific factors and reduce biases associated with treatment, whereas those using WLC, TAU, or delayed treatment provide more moderate levels of causal inference [32,33]. In this review, 10 of the 36 RCTs used AC as a comparator (or included AC in combination), and 5 of the 9 quasi-experimental studies also incorporated ACs. Overall, these findings suggest



XSL•FO
RenderX

that the evidence base for DMHIs targeting older adults is developing unevenly. Although a substantial proportion of studies used RCT designs, the overall quality of evidence is only moderately strong, as many studies used WLC or TAU designs, which limits the strength of causal inference and interpretation.

## Characteristics of the Interventions and Their Clinical Outcomes

More detailed intervention characteristics are provided in Multimedia Appendix 3. CBT was the most commonly used therapeutic approach (n=32), delivered predominantly via websites (26/32) and frequently tailored to meet the needs of older adults. Other therapeutic modalities included mindfulness (n=2), psychoeducation (n=4), and problem-solving therapy (n=2). Less common approaches included life review therapy (n=1), positive psychology (n=1), acceptance and commitment therapy (n=1), cognitive behavioral stress management (n=1), stress inoculation training (n=1), and cognitive restructuring (n=1).

Regarding the targeted mental health conditions, most interventions focused on depression (n=30), followed by general psychological distress (n=11), psychological distress associated with specific physical health conditions (n=12), anxiety (n=8), and comorbid anxiety and depression (n=8). A small number of studies focused on posttraumatic stress disorder (PTSD; n=2) and grief (n=1).

Regarding technology requirements, 16 of the 72 studies required older adults to own a device, 32 provided devices for participants, and 24 did not report this information. This suggests that at least some studies considered technology access and tried to be digitally inclusive. In terms of technology type, the most frequently used was web-based platforms (n=38), followed by games or exergames (n=7), VR (n=6), mobile apps (n=6), and chatbots (n=4). Less commonly used technologies included robots (n=3), video-only programs (n=3), audio-only programs (n=2), voice assistants (n=1), offline platforms (n=1), and wearable devices (n=1).

Regarding the intervention period, 11 studies lasted fewer than 8 weeks, 31 lasted 8 to 12 weeks, 2 extended beyond 12 weeks, and 28 did not clearly report the program duration. In terms of the number of sessions, 52 studies included fewer than 4 sessions, 11 involved 4 to 8 sessions, and 9 had more than 8 sessions. Overall, most interventions were relatively short in duration, clustering within the under-8-week and 8‐to 12-week ranges, while only a few extended into longer-term formats. Interventions with fewer sessions also predominated compared with those involving more intensive or extended session series.

Out of 72 papers reviewed, 54 reported positive clinical outcomes. Among these 54 interventions, the majority—33 (61.11%)—used web-based platforms. Other interventions with positive clinical outcomes included 6 games (11.11%), 5 VR (9.26%), 2 mobile apps (3.70%), 3 chatbots (5.56%), 1 robot (1.85%), 1 offline-platform program (1.85%), 1 video-only program (1.85%), 1 audio-only program (1.85%), and 1 voice assistant (1.85%). These findings show that current DMHIs generally demonstrate benefits to users and that positive outcomes are not limited to a single technology type but are instead distributed across multiple technologies.

Among the 54 interventions that reported positive clinical outcomes, 23 targeted depression as a primary or secondary outcome (42.59%), 8 focused on general psychological distress (14.81%), 8 were designed for psychological distress associated with physical health conditions (14.81%), 6 addressed anxiety (11.11%), 6 targeted both anxiety and depression simultaneously (11.11%), 2 focused on PTSD (3.70%), and 1 on grief (1.85%). These findings highlight that DMHIs have demonstrated effectiveness across a wide spectrum of mental health conditions.

## Therapeutic Approaches, Technological Platforms, and Adaptations for Older Adults

Table 3 presents the distribution of DMHIs across different mental health conditions. Web-based platforms were the most commonly used technology across nearly all conditions (total n=38), particularly for depression (n=17), psychological distress with comorbid physical illness (n=7), and comorbid anxiety and depression (n=6). Game-based interventions were the second-most frequently used (total n=7), targeting depression (n=4), and general psychological distress (n=2). These were followed by VR interventions (total n=6), applied across a more diverse set of mental health issues, including depression (n=2), general psychological distress (n=1), psychological distress with comorbid physical illness (n=1), PTSD (n=1), and grief (n=1). Mobile apps (total n=6) were also distributed across multiple conditions, including depression (n=2), general psychological distress (n=1), psychological distress with comorbid physical illness (n=2), and comorbid anxiety and depression (n=1). Other technologies, such as chatbots, voice assistants, audio-only programs, video-only programs, and robots, appeared in smaller numbers.



XSL•FO
RenderX

**Table .** Cross-tabulation of technology types and mental health conditions.

| | Depression | General psychological distress | Psychological distress with physical condition | Anxiety | Anxiety and depression | Posttraumatic stress disorder | Grief | Total |
|---|---|---|---|---|---|---|---|---|
| Web-based | 17 | 4 | 7 | 3 | 6 | 1 | 0 | 38 |
| Game | 4 | 2 | 0 | 1 | 0 | 0 | 0 | 7 |
| VR[a] | 2 | 1 | 1 | 0 | 0 | 1 | 1 | 6 |
| Mobile app | 2 | 1 | 2 | 0 | 1 | 0 | 0 | 6 |
| Chatbot | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 4 |
| Robot | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 3 |
| Video only | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 3 |
| Audio only; offline platform | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Voice assistant | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Offline platform | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Wearable technology | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Total | 30 | 11 | 12 | 8 | 8 | 2 | 1 | 72 |

[a]VR: virtual reality.

Across intervention types, web-based programs, particularly those using CBT, were the most developed and commonly used. These interventions were often applied to depression, anxiety, and psychological distress and were most frequently associated with positive clinical outcomes. However, the larger number of web-based CBT studies does not necessarily mean that this modality is superior to other technologies. Mobile apps, VR, chatbots, robots, and voice assistants appeared less frequently, but they suggest a growing diversification of DMHIs for older adults and may offer opportunities for more interactive, creative, and engaging forms of intervention delivery. However, studies and evidence on these technologies remain relatively limited.

In reviewing the designed features of DMHIs, this study also found that older adults, as users, often faced difficulties with navigating online platforms, limited internet skills, and technical issues. Common issues reported by users included poor web interface usability, accessibility barriers, and content perceived as nonengaging [34]. Low digital literacy further complicated the use of mobile apps and web-based interventions, as many older adults struggled with smartphone navigation, small screen sizes, and technical problems, which limited sustained engagement [35]. Some participants also struggled to see the personal relevance of the interventions, resulting in low sustained engagement. They expressed a desire for greater customization and improved mobile functionality to enhance relevance [36].

In response, adjustments were made to better tailor interventions to older adults' needs. Adaptations to enhance accessibility, usability, and engagement were reported in 23 of the 72 studies, indicating that adaptations tailored to older adults remain relatively uncommon and present substantial opportunities for future development. Among these 23 studies, usability-related design improvements included larger buttons, high-contrast visuals, simplified language, reduced text, and larger fonts in printed training materials [37]. In addition, several interventions used age-specific cases, scenarios, or narratives in educational materials. For instance, the treatment protocol for depression and anxiety was modified by incorporating age-appropriate strategies [38] such as case examples and skill demonstrations tailored specifically for older adults. Similarly, studies also incorporated assistive devices—including sensor pens, tablet holders, and touchscreen tools—to accommodate physical limitations and enhance the accessibility of information and communications technology—communication interventions and information and communications technology—entertainment interventions [39]. Game-based interventions commonly included features such as slower pace, low stimulation, minimal competition, and calming visuals and sounds to match older adults' physical capacities and processing speeds. For example, 1 study [40] designed cognitive control interventions with these features to reduce anxiety and improve engagement. Moreover, participatory design approaches were commonly used to guide product development to better fit the needs of older adults. For example, 1 study [36] used a "blue-sky" approach—a creative, open-ended design process—to create prototypes focused on meeting the unique needs of older adults.

Specific tailoring strategies varied by technology type (Table 4). Web-based interventions (n=13) most frequently incorporated age-friendly design features, such as large fonts, simplified navigation, and age-relevant case examples. Game-based



XSL•FO
RenderX

interventions (n=4), often used in physical exercise programs, reduced difficulty levels and incorporated calming visuals and gentle physical movements, sometimes supported by physical aids. Mobile apps (n=3) offered simplified interfaces and peer support features, with some specifically developed for long-term care settings to address both cognitive and physical health needs. Audio-only and offline platforms (n=2), along with chatbots and robots (n=1 each), incorporated button-based interfaces and multimodal input options (eg, voice and text) to facilitate ease of use and social interaction.

**Table .** Tailoring of digital mental health interventions for older adults.

| Technology type and target mental health conditions (studies) | Tailoring to the needs of older adults |
|---|---|
| Web-based (n=13) | |
| Depression (n=7) | • Used age-specific cases or scenarios in educational materials.<br>• Applied age-friendly design features to improve usability, such as larger buttons, high-contrast visuals, simplified language, less text, and larger fonts in printed training materials |
| Anxiety and depression (n=2) | • Adjusted content using older adults' narratives and experiences |
| General psychological distress (n=2) | • Incorporated assistive devices, such as sensor pens, tablet holders, and touchscreen tools, to accommodate physical limitations<br>• Used participatory design approaches to tailor product features |
| Psychological distress with physical conditions (n=1) | • Adapted content from interventions previously proven effective for older adults<br>• Modified materials to align with older adults' mental health needs |
| Anxiety (n=1) | • Used age-relevant, real-life scenarios and illustrations<br>• Designed therapeutic modules at a slower pace and with less stress to align with users' processing speed |
| Game (n=4) | |
| Depression (n=2) | • Adjusted game difficulty to match older adults' physical capacities<br>• Incorporated safety-focused designs considering energy and mobility limitations |
| Anxiety (n=1) | • Included game features such as slower pace, low stimulation, minimal competition, and calming visuals and sounds |
| General psychological distress (n=1) | • Incorporated gentle physical activities and encouraged social interaction through gameplay |
| Mobile app (n=3) | |
| Depression (n=2) | • Considered care settings (eg, clinical, community, long-term care) in design<br>• Simplified interfaces to support older adults with cognitive and communicative limitations |
| Psychological distress with physical conditions (n=1) | • Included peer support features targeting both mental and physical health needs of older adults |
| Audio only or offline platform (n=2) | |
| Depression (n=2) | • Based on CBT[a] therapist manuals tailored for older adults<br>• Incorporated age-specific examples, larger font sizes, and simplified navigation to improve accessibility |
| Chatbot (n=1) | |
| Anxiety and depression (n=1) | • Built on a popular platform already familiar to older adults<br>• Used button-based interfaces and age-appropriate exercise videos to improve usability and relevance |

[a]CBT: cognitive behavioral therapy.

In Table 4, there are 4 main types of tailoring strategies synthesized from the included studies: interface adaptations, content adaptations, delivery support, and participatory design. Interface adaptations included larger fonts, simplified



XSL•FO
RenderX

navigation, high-contrast visuals, button-based interfaces, and reduced text. Content adaptations included age-relevant examples, older adult narratives, slower pacing, and scenarios reflecting later-life experiences. Delivery support included technical assistance, coaching, device provision, and assistive tools to reduce barriers related to digital literacy and physical limitations. Finally, there are also a few interventions that incorporated participatory approaches, in which older adults contributed to intervention development, prototype testing, or usability refinement. All 4 types are important, but they have been developed to different degrees. In existing DMHIs, age-centered design is most often implemented through usability modifications, while deeper forms of participatory development and systematic testing of adaptation strategies remain limited.

Although the reviewed studies did not directly test whether age-related adaptations and specific features themselves contributed to clinical improvements, several features appeared to frequently co-occur with positive engagement with the intervention or clinical outcomes. For example, interventions that provided personalized support, such as coaching or technical assistance, often reported higher engagement and adherence. One example is the Empower@Home program, an intervention whose intellectual property is owned by the Regents of the University of Michigan and which was developed by the author XX and her team [37]. Similarly, programs incorporating social interaction components, such as peer support or group activities, were found to reduce loneliness and improve psychiatric self-management [10]. Age-appropriate content, scenarios, and illustrations were also present in several effective interventions [37,38], while interactive, feedback, and motivational elements appeared in programs with both psychosocial and physical benefits [41].

Taken together, these patterns suggest that age-friendly and age-centered design features may enhance the acceptability and potential benefits of DMHIs for older adults. However, because previous studies did not explicitly test the effects of individual design elements—aside from a few qualitative and mixed methods studies that explored these issues [37,42]—future research should examine these features systematically to clarify their specific contributions to engagement and outcomes.

## Discussion

### Summary of Study Findings

This scoping review provides an overview of the current development of DMHIs targeting older adults. These DMHI programs have used a range of technologies, such as web-based platforms, mobile apps, VR, and chatbots. In recent years, the rapid development of AI has offered new possibilities for personalizing and adapting DMHIs even further. However, we identified few studies that used AI. This may partly reflect the limitations of not including specific AI-related terms in the literature search and our general focus on digital interventions, as well as the typical lag between recent technological innovation—such as the large language model boom in industry—its adoption in real-world products, and subsequent academic work [43]. As emerging technologies like AI continue to advance, future research on DMHIs for older adults should incorporate the evolution and upgrading of technology and examine how these exciting changes shape intervention design, delivery, and evaluation, as well as how emerging technologies can be integrated ethically and effectively into DMHIs targeting older adults in real-world practice [44].

This review also informs us that, beyond technological innovation, tailoring products to older adult users' needs and making products aging-appropriate are also important. The active participation of older adults in both the design and evaluation stages can help make products accessible, usable, and age-friendly. Evidence suggests that involving older adults throughout the design process helps ensure that interventions align with users' needs and preferences, which can improve both engagement and effectiveness [45,46]. However, this review found that only a few studies applied a co-design approach to developing or evaluating DMHIs for older adults—for example, a co-design workshop to test a chatbot prototype [47] and a brainstorming session on mobile app design [36]. Future DMHI design or adaptation should consider actively engaging older adults across all stages to help ensure that interventions are culturally and contextually relevant.

Regarding age-adaptation design, in this review, 31.94% (n=23/72) of the included studies developed or tested interventions that incorporated older adult–centered features to enhance usability and engagement, which shows that current DMHIs with older adult–centered design are underdeveloped to some extent. One possible reason is that many interventions are originally designed for the general population and then adapted for older adults during the launch or usability testing stage to expand the user market [48,49]. This indicates an existing gap between the growing realization of the importance of older adults as target users and the development of underlying design principles for age-centered tailoring in intervention development or improvement.

Also, prior research has highlighted the unique challenges that older adults face when using DMHIs, including usability barriers, accessibility issues, and difficulties integrating technology into daily routines [50]. Many of the interventions in these 23 studies incorporated age-friendly design features, such as larger buttons, high-contrast visuals, simplified language, and larger fonts to accommodate physical and cognitive limitations, which aligns with the aim of overcoming the challenges mentioned above. These 23 studies also inform us that most age-related adaptations are simple and focus on improving ease of use and accessibility tailored to older adults, which also echoes previous studies that caution that complex or visually "fancy" interfaces do not necessarily improve outcomes; rather, simple and accessible designs tend to be more usable and acceptable [14,50].

Regarding future research and practice, the review establishes multiple essential research paths that will help scientists and practitioners enhance their understanding of DMHIs for older adults. First, this study revealed that while nearly half of the evaluation studies used RCT designs, most studies used waitlist or TAU control groups as their comparison conditions. These research designs may limit the ability to establish clear cause-and-effect relationships and may also produce exaggerated



XSL•FO
RenderX

treatment outcomes [31]. Future research should strengthen methodological rigor by using RCTs with active comparison groups and longer follow-up periods. In addition, future evaluations should not only assess clinical effectiveness but also examine whether interventions are usable, accessible, and appropriate for older adults in real-world settings.

Second, the studies presented in Tables 2 and 3 demonstrate that most investigations took place in high-income countries, particularly Australia and the United States, while minority groups and people from low- and middle-income countries remained underrepresented. This limited representation restricts the generalizability of the findings and may contribute to digital health disparities, as interventions developed in resource-rich settings may not reflect the needs, preferences, language backgrounds, cultural contexts, or technology access barriers of diverse older adult populations. From a participatory design perspective, the development of inclusive DMHIs requires researchers to actively involve diverse older adults and community stakeholders throughout the design, adaptation, and evaluation process [48]. Future research should, therefore, expand studies in low- and middle-income countries and among diverse older adult populations to better address mental health disparities [51].

Third, this review found that only about one-third of studies incorporated usability adaptations or age-centered design strategies, while most did not sufficiently address digital literacy, technology accessibility, or user interface challenges faced by older adults. From a universal design perspective, DMHIs should aim to reduce barriers related to accessibility, usability, health literacy, sensory changes, and technology confidence, while avoiding the assumption that older adults are a homogeneous user group [52]. Gerontechnology frameworks similarly emphasize the importance of aligning technological design with the physical, cognitive, emotional, and social dimensions of aging [53]. Therefore, future interventions should move beyond simply adapting existing digital tools for older adults and instead use more theory-informed and participatory approaches to age-centered design.

## Limitations

This scoping review has several methodological limitations that should be considered when interpreting the findings. First, our search strategy was intentionally broad to capture the wide range of DMHIs. While this approach allowed us to map the landscape of existing interventions, it may have missed studies in niche areas that use more specialized terminology. For example, we did not include more specific terms such as machine learning, deep learning, natural language processing, large language model, and generative AI. Therefore, our review may have limited the inclusion of AI-powered interventions. As AI-enabled DMHIs may differ from traditional DMHIs in terms of user interaction, data use, data safety, and ethical risks, the lack of inclusion of AI-enabled DMHIs may limit our ability to capture important trends in this rapidly emerging area of DMHI research. Second, as a scoping review, our goal was to describe and categorize the types of technologies, intervention designs, and adaptations rather than to evaluate intervention efficacy quantitatively. Because we included a wide variety of study designs, intervention types, and outcomes, our ability to compare or quantitatively synthesize the study findings was limited. Another limitation is that we did not conduct a formal quality appraisal or risk-of-bias assessment in this study. While formal quality appraisal is not always required for scoping reviews, the absence of such an assessment limits our ability to evaluate the methodological rigor of each study. Because studies with different designs, samples, comparison conditions, and follow-up periods were synthesized together, this study provides a descriptive overview that scopes and maps the current evidence rather than offering evidence of intervention effectiveness or study quality across all included studies. Future reviews with a narrower scope could conduct formal quality appraisal and quantitative synthesis to better assess intervention effectiveness. Third, the quality of individual studies was not systematically assessed. Judging by the range of study designs included, many interventions would benefit from more rigorous evaluation methods, such as RCTs with adequate sample sizes and follow-up periods. In addition, we only included studies published in English, which likely explains why most included studies were conducted in Western or Global North countries. Finally, we primarily included peer-reviewed journal studies and may have missed gray literature, potentially leading to publication bias.

## Conclusions

This review examined the increasing yet unequal growth of DMHIs, which provide services to adults who are 50 years or older. The 72 studies demonstrated that DMHIs using web-based interventions and other digital tools, including games, VR, mobile apps, chatbots, and robots, produced positive results for the treatment of depression, anxiety, and psychological distress in older adults. The current evidence supporting DMHIs for older adults remains moderate because most studies lack active comparison groups and fail to demonstrate real-world implementation. The research found that age-related adaptations and co-design methods to address age-specific design requirements and usage remained limited throughout the study. Future research needs to establish trials with active comparison groups and extended observation periods, as well as conduct studies on DMHI implementation in regular health care services and community-based environments while including diverse older adult groups and testing user-friendly features with technology training and access support to develop better DMHIs for older adults.

## Acknowledgments

ChatGPT (OpenAI, GPT-4) was used to assist with proofreading and language editing of this manuscript.



XSL•FO
RenderX

## Funding

The authors declare that no financial support was received for this work.

## Authors' Contributions

LZ led and designed the study, including developing the search strategy, conducting the literature search, screening studies, extracting and analyzing the data, interpreting the findings, and leading the drafting of the manuscript. XZ contributed to table preparation and manuscript drafting and revision. JC and CZ participated in study screening and data extraction. SJ participated in study screening. XX supervised the study process, guided the study design, resolved conflicts during the review process, and critically revised the manuscript. All authors reviewed and approved the final manuscript before submission.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms used for article retrieval in databases.
[DOCX File, 9 KB - i-jmr_v15i1e92542_app1.docx ]

Multimedia Appendix 2
Descriptive analysis of the 72 included studies.
[DOCX File, 30 KB - i-jmr_v15i1e92542_app2.docx ]

Multimedia Appendix 3
Characteristics of interventions in the 72 included studies.
[DOCX File, 57 KB - i-jmr_v15i1e92542_app3.docx ]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 672 KB - i-jmr_v15i1e92542_app4.pdf ]

## References

1. Reynolds CF, Jeste DV, Sachdev PS, Blazer DG. Mental health care for older adults: recent advances and new directions in clinical practice and research. World Psychiatry 2022 Oct;21(3):336-363. [doi: 10.1002/wps.20996] [Medline: 36073714]
2. Javed A, Lee C, Zakaria H, et al. Reducing the stigma of mental health disorders with a focus on low- and middle-income countries. Asian J Psychiatr 2021 Apr;58:102601. [doi: 10.1016/j.ajp.2021.102601] [Medline: 33611083]
3. Oluyede L, Cochran AL, Wolfe M, Prunkl L, McDonald N. Addressing transportation barriers to health care during the COVID-19 pandemic: perspectives of care coordinators. Transp Res Part A Policy Pract 2022 May;159:157-168. [doi: 10.1016/j.tra.2022.03.010] [Medline: 35283561]
4. Coombs NC, Meriwether WE, Caringi J, Newcomer SR. Barriers to healthcare access among U.S. adults with mental health challenges: a population-based study. SSM Popul Health 2021 Sep;15:100847. [doi: 10.1016/j.ssmph.2021.100847] [Medline: 34179332]
5. Kirkbride JB, Anglin DM, Colman I, et al. The social determinants of mental health and disorder: evidence, prevention and recommendations. World Psychiatry 2024 Feb;23(1):58-90. [doi: 10.1002/wps.21160] [Medline: 38214615]
6. DeSa S, Gebremeskel AT, Omonaiye O, Yaya S. Barriers and facilitators to access mental health services among refugee women in high-income countries: a systematic review. Syst Rev 2022 Apr 6;11(1):62. [doi: 10.1186/s13643-022-01936-1] [Medline: 35387680]
7. Elshaikh U, Sheik R, Saeed RKM, Chivese T, Alsayed Hassan D. Barriers and facilitators of older adults for professional mental health help-seeking: a systematic review. BMC Geriatr 2023 Aug 25;23(1):516. [doi: 10.1186/s12877-023-04229-x] [Medline: 37626290]
8. Pescosolido BA, Halpern-Manners A, Luo L, Perry B. Trends in public stigma of mental illness in the US, 1996-2018. JAMA Netw Open 2021 Dec 1;4(12):e2140202. [doi: 10.1001/jamanetworkopen.2021.40202] [Medline: 34932103]
9. Teo K, Churchill R, Riadi I, Kervin L, Wister AV, Cosco TD. Help-seeking behaviors among older adults: a scoping review. J Appl Gerontol 2022 May;41(5):1500-1510. [doi: 10.1177/07334648211067710] [Medline: 35156428]
10. Fortuna KL, Torous J, Depp CA, et al. A future research agenda for digital geriatric mental healthcare. Am J Geriatr Psychiatry 2019 Nov;27(11):1277-1285. [doi: 10.1016/j.jagp.2019.05.013] [Medline: 31196619]
11. Gratzer D, Torous J, Lam RW, et al. Our digital moment: innovations and opportunities in digital mental health care. Can J Psychiatry 2021 Jan;66(1):5-8. [doi: 10.1177/0706743720937833] [Medline: 32603188]



XSL•FO
RenderX

12. Hollis C, Falconer CJ, Martin JL, et al. Annual research review: digital health interventions for children and young people with mental health problems—a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
13. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013;35(4):332-338. [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
14. Lattie EG, Stiles-Shields C, Graham AK. An overview of and recommendations for more accessible digital mental health services. Nat Rev Psychol 2022 Feb;1(2):87-100. [doi: 10.1038/s44159-021-00003-1] [Medline: 38515434]
15. Muñoz RF. Using evidence-based internet interventions to reduce health disparities worldwide. J Med Internet Res 2010 Dec 17;12(5):e60. [doi: 10.2196/jmir.1463] [Medline: 21169162]
16. Saad A, Bruno D, Camara B, D'Agostino J, Bolea-Alamanac B. Self-directed technology–based therapeutic methods for adult patients receiving mental health services: systematic review. JMIR Ment Health 2021 Nov 26;8(11):e27404. [doi: 10.2196/27404] [Medline: 34842556]
17. Singla DR. Digital mental health interventions: an opportunity to enhance care while preserving human connection. Lancet Reg Health Am 2024 Nov;39:100940. [doi: 10.1016/j.lana.2024.100940] [Medline: 39559429]
18. Hedman E, Ljótsson B, Kaldo V, et al. Effectiveness of internet-based cognitive behaviour therapy for depression in routine psychiatric care. J Affect Disord 2014 Feb;155:49-58. [doi: 10.1016/j.jad.2013.10.023] [Medline: 24238951]
19. Johansson O, Bjärehed J, Andersson G, Carlbring P, Lundh LG. Effectiveness of guided internet–delivered cognitive behavior therapy for depression in routine psychiatry: a randomized controlled trial. Internet Interv 2019 Sep;17:100247. [doi: 10.1016/j.invent.2019.100247] [Medline: 31249791]
20. Titov N, Dear BF, Schwencke G, et al. Transdiagnostic internet treatment for anxiety and depression: a randomised controlled trial. Behav Res Ther 2011 Aug;49(8):441-452. [doi: 10.1016/j.brat.2011.03.007] [Medline: 21679925]
21. Sixsmith A, Horst BR, Simeonov D, Mihailidis A. Older people's use of digital technology during the COVID-19 pandemic. Bull Sci Technol Soc 2022 Jun;42(1-2):19-24. [doi: 10.1177/02704676221094731] [Medline: 38603230]
22. Kayser J, Wang X, Wu Z, Dimoji A, Xiang X. Layperson-facilitated internet-delivered cognitive behavioral therapy for homebound older adults with depression: protocol for a randomized controlled trial. JMIR Res Protoc 2023 Feb 22;12:e44210. [doi: 10.2196/44210] [Medline: 36811937]
23. O'moore KA, Newby JM, Andrews G, et al. Internet cognitive–behavioral therapy for depression in older adults with knee osteoarthritis: a randomized controlled trial. Arthritis Care Res (Hoboken) 2018 Jan;70(1):61-70. [doi: 10.1002/acr.23257] [Medline: 28426917]
24. Read J, Sharpe L, Burton AL, et al. A randomized controlled trial of internet-delivered cognitive behaviour therapy to prevent the development of depressive disorders in older adults with multimorbidity. J Affect Disord 2020 Mar 1;264:464-473. [doi: 10.1016/j.jad.2019.11.077] [Medline: 31767215]
25. Philippe TJ, Sikder N, Jackson A, et al. Digital health interventions for delivery of mental health care: systematic and comprehensive meta-review. JMIR Ment Health 2022 May 12;9(5):e35159. [doi: 10.2196/35159] [Medline: 35551058]
26. Riadi I, Kervin L, Dhillon S, et al. Digital interventions for depression and anxiety in older adults: a systematic review of randomised controlled trials. Lancet Healthy Longev 2022 Aug;3(8):e558-e571. [doi: 10.1016/S2666-7568(22)00121-0] [Medline: 36102765]
27. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
28. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5(1):69. [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
29. Chakrabarti S. Usefulness of telepsychiatry: a critical evaluation of videoconferencing-based approaches. World J Psychiatry 2015 Sep 22;5(3):286-304. [doi: 10.5498/wjp.v5.i3.286] [Medline: 26425443]
30. Sorkin DH, Janio EA, Eikey EV, et al. Rise in use of digital mental health tools and technologies in the United States during the COVID-19 pandemic: survey study. J Med Internet Res 2021 Apr 16;23(4):e26994. [doi: 10.2196/26994] [Medline: 33822737]
31. Harrer M, Cuijpers P, Schuurmans LKJ, et al. Evaluation of randomized controlled trials: a primer and tutorial for mental health researchers. Trials 2023 Aug 30;24(1):562. [doi: 10.1186/s13063-023-07596-3] [Medline: 37649083]
32. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology 2010 May;21(3):383-388. [doi: 10.1097/EDE.0b013e3181d61eeb] [Medline: 20335814]
33. Cunningham JA, Kypri K, McCambridge J. Exploratory randomized controlled trial evaluating the impact of a waiting list control design. BMC Med Res Methodol 2013 Dec 6;13(1):150. [doi: 10.1186/1471-2288-13-150] [Medline: 24314204]
34. Xiang X, Kayser J, Sun Y, Himle J. Internet-based psychotherapy intervention for depression among older adults receiving home care: qualitative study of participants' experiences. JMIR Aging 2021 Nov 22;4(4):e27630. [doi: 10.2196/27630] [Medline: 34813491]
35. Similä H, Immonen M, Toska-Tervola J, et al. Feasibility of mobile mental wellness training for older adults. Geriatr Nurs 2018;39(5):499-505. [doi: 10.1016/j.gerinurse.2018.02.001] [Medline: 29530292]




XSL•FO
RenderX

36. LaMonica HM, Roberts AE, Davenport TA, Hickie IB. Evaluation of the usability and acceptability of the InnoWell platform as rated by older adults: survey study. JMIR Aging 2021 Apr 21;4(2):e25928. [doi: 10.2196/25928] [Medline: 33881410]
37. Xiang X, Kayser J, Turner S, Ash S, Himle JA. Layperson-supported, web-delivered cognitive behavioral therapy for depression in older adults: randomized controlled trial. J Med Internet Res 2024 Mar 4;26:e53001. [doi: 10.2196/53001] [Medline: 38437013]
38. Dear BF, Zou J, Titov N, et al. Internet-delivered cognitive behavioural therapy for depression: a feasibility open trial for older adults. Aust N Z J Psychiatry 2013 Feb;47(2):169-176. [doi: 10.1177/0004867412466154] [Medline: 23152358]
39. Chiu CJ, Wu CH. Information and communications technology as a health promotion method for older adults in assisted-living facilities: three-arm group-randomized trial. JMIR Aging 2019 May 6;2(1):e12633. [doi: 10.2196/12633] [Medline: 31518261]
40. Anguera JA, Gunning FM, Areán PA. Improving late life depression and cognitive control through the use of therapeutic video game technology: a proof-of-concept randomized trial. Depress Anxiety 2017 Jun;34(6):508-517. [doi: 10.1002/da.22588] [Medline: 28052513]
41. Chao YY, Scherer YK, Montgomery CA, Wu YW, Lucke KT. Physical and psychosocial effects of Wii Fit exergames use in assisted living residents: a pilot study. Clin Nurs Res 2015 Dec;24(6):589-603. [doi: 10.1177/1054773814562880] [Medline: 25488422]
42. Xiang X, Zhang X, Narbut E, et al. Internet-based cognitive behavioral therapy with layperson support: a mixed-methods evaluation of Empower@ Home. Clin Gerontol 2025 Jul 2;2025:1-15. [doi: 10.1080/07317115.2025.2528238] [Medline: 40601411]
43. Comin D, Mestieri M. Technology diffusion: measurement, causes, and consequences. In: Handbook of Economic Growth: Elsevier; 2014:565-622. [doi: 10.1016/B978-0-444-53540-5.00002-1]
44. Löchner J, Carlbring P, Schuller B, Torous J, Sander LB. Digital interventions in mental health: an overview and future perspectives. Internet Interv 2025 Jun;40:100824. [doi: 10.1016/j.invent.2025.100824] [Medline: 40330743]
45. Easton K, Potter S, Bec R, et al. A virtual agent to support individuals living with physical and mental comorbidities: co-design and acceptability testing. J Med Internet Res 2019 May 30;21(5):e12996. [doi: 10.2196/12996] [Medline: 31148545]
46. Darley A, Carroll Á. Conducting co-design with older people in a digital setting: methodological reflections and recommendations. Int J Integr Care 2022;22(4):18. [doi: 10.5334/ijic.6546] [Medline: 36561633]
47. Kilfoy A, Hsu TCC, Stockton-Powdrell C, Whelan P, Chu CH, Jibb L. An umbrella review on how digital health intervention co-design is conducted and described. NPJ Digit Med 2024 Dec 23;7(1):374. [doi: 10.1038/s41746-024-01385-1] [Medline: 39715947]
48. Yin R, Martinengo L, Smith HE, Subramaniam M, Griva K, Tudor Car L. The views and experiences of older adults regarding digital mental health interventions: a systematic review of qualitative studies. Lancet Healthy Longev 2024 Nov;5(11):100638. [doi: 10.1016/j.lanhl.2024.08.007] [Medline: 39491881]
49. Dadich A, Lan A, Shanmugarajan S, Childs S, Alford J, Chróinín DN. Models of care for older people: a scoping review. J Am Geriatr Soc 2025 May;73(5):1588-1597. [doi: 10.1111/jgs.19371] [Medline: 39831543]
50. Ying Y, Ji Y, Kong F, et al. Internet-based cognitive behavioral therapy for psychological distress in older adults without cognitive impairment living in nursing homes during the COVID-19 pandemic: a feasibility study. Internet Interv 2021 Dec;26:100461. [doi: 10.1016/j.invent.2021.100461] [Medline: 34631432]
51. Sengpiel M, Volkmann T, Jochems N. Considering older adults throughout the development process: the HCD+ approach. Presented at: Technology for an Ageing Society: Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2018 Annual Conference; Oct 10-12, 2018 URL: https://www.hfes-europe.org/wp-content/uploads/2018/10/Sengpiel2018.pdf [accessed 2026-06-09]
52. Farage MA, Miller KW, Ajayi F, Hutchins D. Design principles to accommodate older adults. Glob J Health Sci 2012 Feb 29;4(2):2-25. [doi: 10.5539/gjhs.v4n2p2] [Medline: 22980147]
53. Fozard JL, Rietsema J, Bouma H, Graafmans JAM. Gerontechnology: creating enabling environments for the challenges and opportunities of aging. Educ Gerontol 2000 Jun;26(4):331-344. [doi: 10.1080/036012700407820]


## Abbreviations

**AC:** active comparator
**AI:** artificial intelligence
**CBT:** cognitive behavioral therapy
**DMHI:** digital mental health intervention
**PICO:** population, intervention, comparator, and outcome
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
**PTSD:** posttraumatic stress disorder



XSL•FO
RenderX

**RCT:** randomized controlled trials
**TAU:** treatment as usual
**VR:** virtual reality
**WLC:** waitlist control

*Edited by M Balcarras; submitted 30.Jan.2026; peer-reviewed by L Ilodigwe, V Yadav; revised version received 12.May.2026; accepted 01.Jun.2026; published 06.Jul.2026.*

*Please cite as:*
*Zhou L, Zuo X, Cho J, Zheng C, Jing S, Xiang X*
*Landscape, Evidence, Gaps, and Opportunities in Digital Mental Health Interventions for Older Adults: Scoping Review*
*Interact J Med Res 2026;15:e92542*
*URL: https://www.i-jmr.org/2026/1/e92542*
*doi:10.2196/92542*

© Leyi Zhou, Xinyi Zuo, Joonyoung Cho, Chuxuan Zheng, Shengyi Jing, Xiaoling Xiang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 6.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Trust Barriers and Vulnerabilities in Older Adults' Telemedicine Adoption: Scoping Review

Wanqing Wang[1]; Xin Ma[1]; Matthew Ting Chi Liu[2]; Angela Chang[1]

[1]Department of Communication, Faculty of Social Sciences, University of Macau, Avenida da Universidade, Taipa, Macau, China Special Administrative Region

[2]Department of Management and Marketing, Faculty of Business Administration, University of Macau, Macau, China

**Corresponding Author:**
Angela Chang
Department of Communication, Faculty of Social Sciences, University of Macau, Avenida da Universidade, Taipa, Macau, China Special Administrative Region

## Abstract

**Background:** Aging populations worldwide face increasing health care demands, particularly for chronic disease management. While telemedicine offers a viable solution to enhance health care access, significant trust-related barriers hinder its adoption among older adults, warranting a scoping synthesis of available evidence.

**Objective:** This scoping review aims to map the available evidence on trust barriers experienced by older adults in telemedicine, identify underlying vulnerability domains, and chart the evidence base for design and policy recommendations.

**Methods:** We conducted a scoping review in accordance with PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines, analyzing literature from PubMed, Web of Science, and Scopus. Thematic analysis was applied to synthesize findings from 30 included studies.

**Results:** Four primary trust barriers were identified: technophobia and technical difficulties, privacy and data security concerns, negative emotional and social impacts, and a strong preference for in-person care. These barriers mapped onto 4 vulnerability domains: limited telemedicine literacy (particularly low eHealth self-efficacy), declining health status (including sensory and cognitive impairments), psychological and cognitive factors (such as anxiety about losing autonomy), and inadequate social support systems. The review also underscored how rapid technological change amplifies these challenges for older adults.

**Conclusions:** Effective telemedicine implementation for older adults requires multipronged interventions, including age-appropriate interface design, targeted digital literacy training, robust privacy protections, and personalized support systems. These approaches address both technological and psychosocial barriers, potentially increasing engagement while mitigating vulnerabilities. Future research should assess the effectiveness of these interventions across diverse older populations.

***(Interact J Med Res 2026;15:e84818)*** doi:10.2196/84818

**KEYWORDS**

trust; vulnerability; telemedicine; older adults; digital health; health equity

## Introduction

### Background

As demographic transitions advance, characterized by declining fertility and increasing mortality in most countries, along with increased life expectancy due to the decline in deadly infectious diseases, there has been a remarkable increase in global population aging. It is estimated that by 2050, the worldwide aging rate will rise to 15.9%, with projections indicating it could reach 22.4% by the year 2100 [1,2]. Aging populations present a growing health care challenge, as older adults often require care for chronic diseases such as heart conditions, diabetes, or respiratory issues. Traditional in-person health care, hindered by mobility limitations, transportation difficulties, and frequent appointments, is often inaccessible, while telemedicine offers a promising alternative to enhance convenience [3-5].

Telemedicine can be broadly defined as the use of telecommunication technologies to provide medical information and services [6-9]. In practice, it encompasses a suite of resources designed to facilitate communication between patients and health care providers, including online consultations, remote monitoring, and virtual rehabilitation. Nevertheless, addressing trust-related challenges is essential for the widespread adoption of telemedicine, particularly among older adult populations [7,10-12].

Trust is commonly defined as the willingness of trustors to accept vulnerability, and the level of vulnerability at the individual level is key to understanding trust [13-15]. Perceived trust issues have been documented in existing research among



XSL•FO
RenderX

older adults, including the use of digital health technologies [9,16,17], concerns about information privacy [18,19], and doctor-patient interaction [14-16]. Trust deficits among older adults have been identified as substantive obstacles to telemedicine uptake, limiting service acceptance and access to health care [20,21].

While prior studies have highlighted the trust issues that older adults encounter in telemedicine, a critical knowledge gap remains. Although trust is commonly defined as an acceptance of vulnerability [13-15], more research has emphasized surface-level factors underlying trust concerns. The deeper dimensions of vulnerability that directly shape trust perceptions remain underexplored. Because vulnerability is not simply an outcome but an intrinsic element of trust dynamics, it cannot be overlooked. In this review, we conceptually distinguish trust barriers as the observable manifestations of distrust or resistance in the context of telemedicine, and vulnerabilities as the underlying predisposing states or conditions that create susceptibility to these barriers. This distinction follows Hamm et al [13], who define vulnerability as an intrinsic antecedent of trust dynamics rather than an outcome. Without clarifying how specific vulnerabilities manifest and interact among older adults, strategies to enhance trust in telemedicine risk lacking the precision necessary to dismantle entrenched barriers and improve health care access for this population.

### Objective

This scoping review aims to map the available evidence on trust barriers in older adults' telemedicine use; chart how these barriers link to underlying vulnerability dimensions, and as an exploratory extension; and identify practice-informed design recommendations. At present, in the field of gerontechnology and digital health, it remains unclear how trust-eroding factors in telemedicine are interrelated with the vulnerabilities of older adults at the individual, interpersonal, and system levels.

To achieve this, the study focuses on 4 interconnected research questions (RQs), informed by established digital health frameworks (eg, the Technology Acceptance Model for Mobile Health) [22,23] and extended to center on dimensions of vulnerability:

- RQ1: What evidence exists regarding the adoption of telemedicine among older adults in existing research?
- RQ2: What trust-eroding factors in telemedicine are documented among older adults, and how do they vary across individual, interpersonal, and systemic levels?
- RQ3: How do these trust-eroding factors correlate with specific dimensions of vulnerability at the individual, interpersonal, and systematic levels?
- RQ4: What intervention strategies can address these vulnerabilities and enhance trust in telemedicine?

These questions operationalize a multilevel analytical lens, encompassing the individual level (eg, declining health status, sensory impairments, and limited digital literacy) [24,25]; the interpersonal level (eg, patient-provider communication patterns [26,27], caregiver support, and peer influences on technology perceptions); the systemic level (eg, data privacy regulations, platform design standards, and the availability of technical support infrastructure) [24,25,28-30]; and the cumulative interactions across levels [31,32].

By synthesizing the current state of knowledge, this study aims to advance theoretical understanding of how vulnerability shapes trust dynamics in aging populations. It moves beyond surface-level trust factors to examine their root causes within layered vulnerabilities. Practically, it seeks to inform scalable, evidence-based interventions that enhance telemedicine acceptance and effectiveness, bridging empirical insights with real-world solutions.

## Methods

### Overview

A scoping review, combined with thematic analysis, was conducted to examine trust barriers experienced by older adults in telemedicine and their associated vulnerabilities. Scoping reviews are valuable for rapidly mapping key concepts, primary sources, and evidence types within a research area, particularly in complex or underexplored fields [33,34]. This approach is especially effective for (1) identifying the scope of available literature on a specific topic, (2) providing an overview of key concepts, and (3) identifying gaps in the research landscape [35-39]. In contrast, systematic reviews primarily synthesize quantitative outcomes to assess intervention effectiveness [38-40].

Given the fragmented and limited research on the determinants of trust and associated vulnerabilities in telemedicine adoption among older adults, a scoping review with thematic analysis was selected as the methodological approach. Existing studies have largely emphasized surface-level trust issues without systematically exploring the underlying dimensions of vulnerability that shape these perceptions. This omission has left a significant gap in understanding how technological, informational, and relational vulnerabilities interact to influence trust in telemedicine. To address this gap, the combined scoping review and thematic analysis was deemed most appropriate.

Thematic analysis was employed to analyze the qualitative data extracted from the included studies. This approach offers a flexible yet rigorous method for analyzing textual data, enabling the identification, analysis, and reporting of patterns (themes) within the data [40,41]. Thematic analysis was chosen to provide a rich, nuanced understanding of the factors influencing older adults' trust in telemedicine. Examples of studies employing a combined scoping review and thematic analysis approach exist in diverse fields such as medicine, education, and sociology [42-45]. Thematic analysis followed an inductive approach [46], proceeding in 6 stages: familiarization with data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report. Initial coding was performed by the lead author. To establish interrater reliability, the second author independently coded approximately 30% of the extracted excerpts. Interrater agreement was substantial (Cohen $\kappa$=0.78) [47]. Discrepancies were resolved through consensus discussion with all of the coauthors. Thematic saturation was considered achieved when the final 5 studies reviewed yielded no new codes.



XSL•FO
RenderX

The study was conducted and reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines (Checklist 1), an evidence-based framework consisting of a 4-phase flow diagram and a 27-item checklist to ensure methodological rigor and transparency in scoping reviews [48-50].

### Searching and Screening Process

Electronic searches for this scoping review were conducted in May 2025, timed with the World Health Organization (WHO) Global Strategy on Digital Health 2020 to 2025, post its mid-2025 progress review to capture up-to-date evidence on older adults' digital health. Three significant databases were considered as primary sources, namely PubMed, Web of Science, and Scopus, as they include subjects such as health, aging, medicine, digital technology, communication, and psychology.

Before finalizing the search strategy, we explored the 3 databases using key concepts such as "telehealth," "trust," and "vulnerability." Keywords were expanded based on the initial search results. Next, a unified search term incorporating Boolean operators was utilized across all databases:

((((trust AND (vulnerability OR vulnerable)) AND (telemedicine OR telehealth OR eHealth OR mHealth)) AND (elderly OR older adults OR old people)) AND patient).

In PubMed, searches were conducted using both free-text keywords and MeSH terms (eg, "Telemedicine" [MeSH], "Aged" [MeSH], and "Trust" [MeSH]) to ensure comprehensive coverage. The search covered all publication years and was limited to English-language sources. To ensure methodological rigor, only peer-reviewed journal papers and review papers were included.

Inclusion criteria were as follows: (1) studies must directly address trust issues or the vulnerability of old patients in telemedicine; (2) study participants must be aged 55 years or older, or include a dedicated analysis of a group over 55 years old. Reviews with unreported demographics were included only if 75% or more of their synthesized studies explicitly focused on older adults (eg, through keywords such as "geriatric" or "aging"); and (3) studies must be patient-centered and conducted from the patients' perspective.

Exclusion criteria included (1) non–peer-reviewed and nonscholarly materials, such as conference papers, book chapters, reports, design work, personal interview reports, newspaper articles, and other nonacademic publications; (2) studies with missing data or requiring paid access beyond institutional subscriptions, due to practical accessibility limitations; and (3) on PubMed, papers not based on human subjects were specifically excluded. These criteria ensure that the review's findings directly inform its research questions regarding enhancing trust in telemedicine among older adults.

Overall, the search yielded 740 publications across 3 databases, which was reduced to 723 after duplicates were removed using EndNote 21.4 (Clarivate Analytics). The first author independently screened titles and abstracts for relevance, followed by a second round of eligibility assessment. Of these, 253 studies met the inclusion criteria and proceeded to full-text review.

Papers that did not focus on older patients aged 55 years or older or did not discuss issues from the patients' perspective were excluded (86/723, 11.9%), as were papers that did not capture trust issues or the vulnerability of older patients in telemedicine (57/723, 7.9%). In addition, we discarded papers that only briefly mentioned trust and vulnerability without exploring these concepts in depth (80/723, 11.1%). Specifically, studies that referenced trust and vulnerability in passing—such as in the introduction or discussion—without providing a focused analysis, dedicated sections, or empirical evidence directly addressing these concepts were excluded. This decision was made by assessing the relevance of trust and vulnerability to the study objectives, methodologies, and findings.

Data extraction was performed using a standardized charting form developed iteratively. The form was piloted on 5 randomly selected included studies and refined by the research team to ensure completeness and consistency. The final charting form captured the following variables for each included study: (1) author(s) and year of publication, (2) journal and study design, (3) participant age range and sample size, (4) geographic focus and country, (5) telemedicine modality, (6) specific trust barriers identified, and (7) vulnerability dimensions reported. Two independent reviewers performed data extraction, and discrepancies were resolved through discussion or consultation with a third reviewer. Extracted data were cross-verified by both reviewers.

Findings were synthesized using inductive thematic analysis, following the six-phase framework described by Braun and Clarke [46]: (1) familiarization with the data through repeated reading of the extracted text, (2) generating initial codes applied to trust-barrier and vulnerability-related content, (3) searching for themes by grouping related codes, (4) reviewing themes for coherence and distinctiveness, (5) defining and naming final themes, and (6) producing the synthesis report. Codes and emerging themes were discussed among coauthors to enhance interpretive rigor.

## Results

### Characteristics of Included Studies

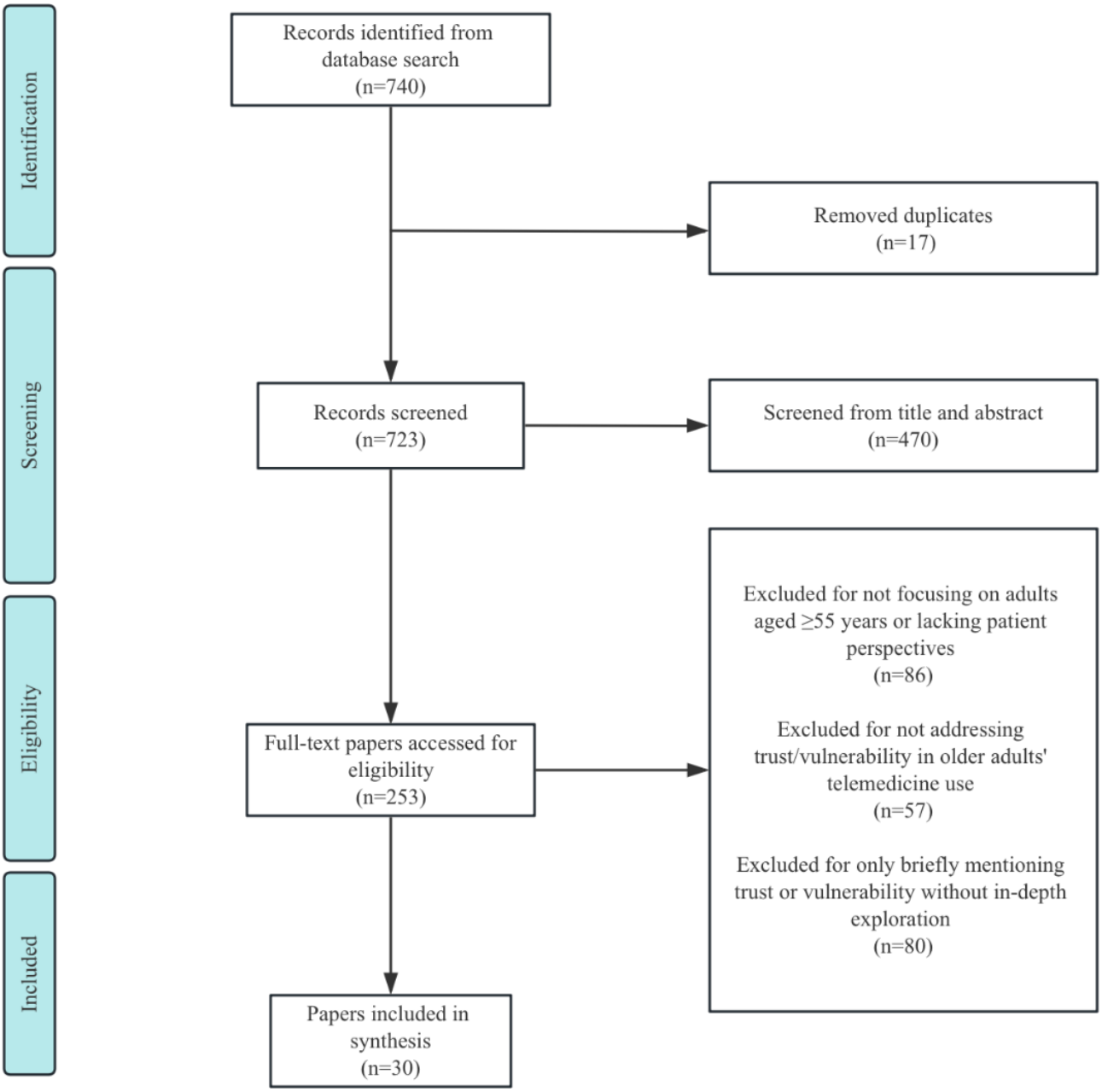
The final sample comprised 30 studies, of which 28 (93.3%) were published between 2020 and 2025 and only 2 (6.7%) prior to 2020. This dominance of post-2020 publications reflects the rapid expansion of telemedicine research during the COVID-19 pandemic, as the public health crisis heightened the need for remote health care solutions. The included studies consisted of quantitative research (11/30, 36.7%), qualitative research (8/30, 26.7%), mixed-methods research (1/30, 3.3%), and reviews (10/30, 33.3%). The 11 quantitative studies (typically large cross-sectional surveys using validated scales) provide the strongest evidence for the relative prevalence of specific barriers. The 8 qualitative studies offer contextually rich descriptions of lived trust experiences, capturing mechanisms that quantitative measures may not fully detect. The 10 review papers



XSL•FO
RenderX

(comprising 5 scoping reviews, 2 systematic reviews, 1 rapid review, 1 realist review, and 1 narrative review) synthesize broader evidence but introduce additional heterogeneity across populations and contexts. The descriptive characteristics of all included studies are presented in Multimedia Appendix 1. A total of 30 papers were included in the review, and the extraction procedures are presented in Figure 1.

**Figure 1.** PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flow diagram of the study selection process for the scoping review.




## Geographic and Economic Distribution of Studies

The included studies were conducted across 12 countries, with China contributing the largest number (7/30, 23.3%), followed by the United Kingdom (3/30, 10%) and the United States (2/30, 6.7%). Other contributing countries included Germany, Sweden, the Philippines, Malaysia, India, Australia, New Zealand, and Finland (each with 1/30, 3.3%). Geographic coverage revealed a stark imbalance: most studies (28/30, 93.3%) focused on high-income regions (Europe, North America, and parts of Asia), with very few addressing low- and middle-income countries, underscoring a global equity gap.

Population size varied by study design. Quantitative studies had larger samples (111 - 1852 participants), whereas qualitative studies (including mixed-methods research) had smaller samples (6 - 98 participants). For review papers, the number of included studies ranged from 10 to 193, except for 1 narrative review that did not report a specific sample size.

Although the general threshold for older adults is often 65 years or more (WHO standard: ≥60), more than a quarter of the included studies (8/30, 26.7%) expanded this range. Specifically, 18 (60.0%) papers reported an age range of 60 years old and above, whereas 8 (26.7%) papers broadened the range to 55 years old and above; 2 (6.7%) papers did not report precise age



XSL•FO
RenderX

ranges (one reported an average age of 66.94 years, and the other reported the majority were over 65 years). Another 2 (6.7%) reviews did not report the age range but were also retained after verifying that over 75% (22/30) of their original studies focused on the older adult population. Researchers cited several reasons for the expansion: (1) 50 or more is a common threshold in technology-use literature, reflecting demographic trends; (2) widening the range avoided excluding high-quality studies and supported inclusive evidence; (3) in some populations (eg, the homeless), 50 or more is recognized as older due to “accelerated aging;” and (4) alignment with the global minimum retirement age of 50 years, which accounts for socioeconomic dimensions of aging.

Across the 30 included studies, telemedicine modalities varied, with 5 common types identified: telemedicine apps (11 papers), wearable devices (10 papers), video consultations (5 papers), electronic surveillance (4 papers), and government services (4 papers). Other modalities included social media services, family therapy, virtual diagnostics, monitoring sensors, SMS text messaging, and online counseling (each reported in 1‐2 papers). Six papers did not specify the modality. Table 1 provides the detailed evidence on older adults’ telemedicine adoption.



XSL•FO
RenderX

**Table .** Characteristics of older adults' telemedicine adoption[a].

| Number | Authors and year | Methods (study design) | Participants' age (y) | Sample size | | Geographic focus (country) | Telemedicine modality |
|---|---|---|---|---|---|---|---|
| | | | | Empirical research sample | Review included studies | | |
| 1 | Essén [51] (2008) | Qualitative research | ≥68 | 17 | —[b] | Sweden | Electronic surveillance |
| 2 | Fitzsimmons et al [52] (2016) | Qualitative research | Mean=66.94 | 9 | — | The United Kingdom | Wearable devices |
| 3 | Kruse et al [53] (2020) | Systematic review | ≥50 | — | 57 | Europe and America | Video consultations, electronic surveillance, and telemedicine apps |
| 4 | Ladin et al [54] (2021) | Qualitative research | ≥70 | 30 | — | The United States | No specified |
| 5 | Schorr et al [55] (2021) | Systemic review | ≥65 | — | 26 | Europe, North America, and Asia | Telemedicine apps, wearable devices, and SMS text messaging |
| 6 | Wilson et al [25] (2021) | Scoping review | ≥60 | — | 14 | Europe, America, and Australia | Telemedicine apps and online counseling |
| 7 | Chen and Liu [56] (2022) | Quantitative research | ≥50 | 390 | — | China | Wearable devices |
| 8 | Hunter et al [57] (2022) | Qualitative research | ≥55 | 98 | — | New Zealand | No specified |
| 9 | Kaihlanen et al [21] (2022) | Qualitative research | ≥65 | 16 | — | Finland | Government service |
| 10 | Singh et al [58] (2022) | Quantitative research | ≥60 | 534 | — | India | Wearable devices |
| 11 | Korkmaz Yaylagul et al [59] (2022) | Scoping review | ≥65 | — | 40 | Europe, North America, and Asia | Video consultations, electronic surveillance, telemedicine apps, and wearable devices |
| 12 | Guo et al [17] (2023) | Quantitative research | ≥50 | 111 | — | China | Telemedicine apps |
| 13 | Liu et al [11] (2023) | Quantitative research | ≥65 | 201 | — | China | Telemedicine apps |
| 14 | Orzechowski et al [60] (2023) | Qualitative research | ≥73 | 11 | — | Germany | Wearable devices |
| 15 | Ali et al [61] (2024) | Qualitative research | ≥65 | 6 | — | The United Kingdom | Telemedicine apps |
| 16 | Chen et al [9] (2024) | Quantitative research | ≥60 | 309 | — | China | No specified |
| 17 | Chu et al [62] (2024) | Quantitative and qualitative research | Survey: 68 (SD 11); Interview: 65 (SD 16) | 55 | — | Australia | Family therapy |
| 18 | Fang et al [63] (2024) | Quantitative research | ≥60 | 277 | — | China | Virtual diagnostics, telemedicine apps, and wearable devices |
| 19 | Gallardo et al [7] (2024) | Quantitative research | ≥60 | 180 | — | The Philippines | Telemedicine apps |



XSL•FO
RenderX

| Number | Authors and year | Methods (study design) | Participants' age (y) | Sample size | | Geographic focus (country) | Telemedicine modality |
|---|---|---|---|---|---|---|---|
| | | | | Empirical research sample | Review included studies | | |
| 20 | Li et al [2] (2024) | Quantitative research | ≥60 | 400 | — | China | No specified |
| 21 | Niu et al [64] (2024) | Quantitative research | ≥60 | 661 | — | China | Government service |
| 22 | Turcotte et al [65] (2024) | Rapid review | ≥50 | — | 24 | Europe, North America, and Asia | Telemedicine apps, video consultations, and wearable devices |
| 23 | Windle et al [66] (2024) | Scoping review | Most≥65 | — | 193 | Europe, North America, and Australia | No specified |
| 24 | Husain and Greenhalgh [67] (2025) | Qualitative research | ≥65 | 17 | — | The United Kingdom | Video consultations |
| 25 | Tan et al [68] (2025) | Quantitative research | ≥60 | 119 | — | Malaysia | Video consultations, wearable devices, and government service |
| 26 | Zhu et al [69] (2025) | Quantitative research | ≥60 | 1852 | — | The United States | No specified |
| 27 | Adams et al [70] (2025) | Scoping review | ≥50 | 10 | — | America | Telemedicine apps and government service |
| 28 | Fothergill et al [30] (2025) | Realist review | Unreported | — | 10 | Europe, America, and Australia | Pendant alarms, monitoring sensors, wearable devices, and electronic surveillance |
| 29 | Western et al [29] (2025) | Narrative review | Unreported | — | Unreported | Global | Social media service, telemedicine apps, and wearable devices |
| 30 | Zhang et al [71] (2025) | Systematic review | ≥50 | — | 16 | Asia and North America | Social media service and telemedicine apps |

[a]Two reviews lacked explicit reporting but were retained after verifying that 75% or more of their primary studies targeted older populations.

[b]Not applicable.

## Dimensions of Trust Barriers

Analysis of the included studies identified 4 thematic categories of trust barriers: technophobia and difficulties, preference for in-person care, privacy and data concerns, and emotional and social impacts.

### *Technophobia and Difficulties*

Technical issues were the most prevalent barrier, reported in 28 (93.3%) papers. Subthemes included low technical skills (eg, inability to navigate interfaces and limited understanding of monitoring mechanisms; 19/30, 63.3%), lack of technical support (eg, absence of family assistance and inadequate troubleshooting services; 14/30, 46.7%), age-related sensory or cognitive decline (eg, impaired vision or hearing and reduced attention span; 12/30, 40.0%), and technical anxiety or fear of using electronic devices (9/30, 30.0%). Less common but notable factors included physical coordination limitations (eg, motor impairments affecting device use; 4/30, 13.3%) and inconsistent user interfaces across platforms (2/30, 6.7%). Modality-specific patterns were evident: studies involving telemedicine applications (11/30, 36.7%) most frequently identified barriers related to interface complexity and low self-efficacy, with participants reporting difficulty navigating functions, performing self-monitoring tasks, or evaluating information reliability. Studies involving wearable devices (10/30, 33.3%) emphasized physical operation difficulties and doubts about data accuracy, reflecting the additional dexterity



XSL•FO
RenderX

and technical comprehension demands of sensor-based technologies.

### *Preference for In-Person Care*

A strong preference for in-person care was another significant barrier, identified in 26 (86.7%) papers. Specifically, 18 (60.0%) papers directly reported a preference for in-person care; 15 (50.0%) noted concerns about the accuracy of telehealth diagnostics or equipment; and 11 (36.7%) indicated reliance on manual verification processes. Low health literacy, contributing to skepticism toward telemedicine, was reported in 5 (16.7%) papers. Cultural influences were highlighted in 3 papers, pointing to disparities in the acceptance of innovative medical approaches across Eastern and Western contexts or between ethnic groups. Studies (4/30, 13.3%) examining government digital health services additionally identified complex digital pathways, inadequate multilingual support, and structural access inequities as barriers reinforcing this preference, suggesting that system-level design failures can compound individual reluctance toward remote care.

### *Privacy and Data Concern*

Privacy and data issues were identified in 23 (76.7%) papers. Concerns included fear of data leakage or breaches (16/30, 53.3%), perceived intrusiveness of continuous monitoring (10/30, 33.3%), and lack of ownership or control over data, including uncertainty about sharing practices (5/30, 16.7%). In a few cases, lack of personal devices or internet access contributed to privacy-related distrust (3/30, 10.0%). Studies involving electronic surveillance technologies (4/30, 13.3%) generated the most acute privacy concerns, encompassing fears about unauthorized third-party access, stigma associated with being continuously monitored, and the perceived intrusiveness of ambient sensors tracking behavioral patterns. Studies involving wearable devices similarly surfaced a concern specific to passive, always-on data collection. However, an exception was noted in a Swedish study on electronic care surveillance, where most participants (16/17) viewed the service positively, citing safety benefits, while only 1 participant expressed discomfort with constant monitoring.

### *Emotional and Social Impacts*

Emotional and social impacts were identified in 21 (70%) papers. Reported issues included discomfort in virtual settings, such as receiving distressing news without perceiving empathy or lacking nonverbal cues (6/30, 20.0%). Concerns about losing personal connections with clinicians were expressed in 12 (40.0%) papers, and refusal to engage due to stigma or embarrassment—such as perceiving telemedicine use as a sign of dependence or incompetence—was noted in 7 (23.3%) papers. Specifically, in 1 (3.3%) mental health study, participants expressed uncertainty about the identity of the remote provider, which heightened distrust. A few secondary factors (eg, financial costs for users and organizational resource constraints in rural areas) were also noted but lacked sufficient empirical support for generalizable conclusions. This barrier was most documented in studies focusing on video consultations (5/30, 16.7%). In synchronous video communication, participants described an intensified sense of detachment resulting from screen-mediated interaction and a reduction in nonverbal cues. In contrast, emotional concerns arising from asynchronous communication primarily centered on a lack of human warmth and continuity.

## Dimensions of Vulnerabilities

Across the 30 included studies, 4 primary categories of vulnerabilities related to telemedicine adoption among older adults were identified: low telemedicine literacy, dependence on social support, psychological and cognitive dispositions, and complex health conditions.

### *Low Telemedicine Literacy*

Low telemedicine literacy was the most frequently documented vulnerability, identified in 29 (96.7%) papers. Three dominant subthemes emerged within this category. Limited knowledge of telehealth devices and services was reported in 22 (73.3%) studies, including insufficient understanding of device functions, data collection mechanisms, and service scope. Dependence on external guidance was observed in 16 (53.3%) studies, with participants requiring assistance from family members, caregivers, or clinical staff to operate devices, navigate platforms, or interpret data. Low self-management skills were noted in 14 (46.7%) studies, reflected in difficulties independently monitoring personal health indicators (eg, blood glucose and blood pressure) or adjusting health behaviors based on remote feedback.

Exceptional cases included a lack of access to telemedicine training resources (eg, absence of community workshops and insufficient user manuals), as reported in 4 (13.3%) studies. Additionally, 3 (10%) papers highlighted language barriers, such as limited English proficiency or the inability to describe symptoms without relying on nonverbal communication.

### *Dependence on Social Support*

Dependence on social support was as prevalent as low telemedicine literacy, identified in 29 (96.7%) studies. This dependency manifests in 3 aspects. Reliance on family members or caregivers for device operation was reported in 21 (70.0%) studies. Participants required assistance with device setup (eg, pairing wearables with smartphones), troubleshooting technical issues, or entering health data. Dependence on clinical staff for health management was noted in 17 (56.7%) studies, encompassing reliance on clinicians for medication adjustments, interpretation of telehealth data, and guidance on disease management through remote channels. Living alone with limited immediate support was observed in 14 (46.7%) studies, where solitary living arrangements heightened concerns about delayed assistance during acute health events (eg, falls and sudden symptom exacerbations) when using telemedicine. Three (10.0%) papers suggested that older adults may lack intrinsic health behavioral drivers and rely more on external incentives to motivate participation in health-related activities.

Apart from artificial assistance, 5 (16.7%) papers reported financial dependency as a barrier to telemedicine access, such as the inability to afford devices or internet services. Additionally, 3 papers indicated difficulties accessing general practitioners due to structural barriers (eg, appointment shortages and geographic inaccessibility).



XSL•FO
RenderX

### *Psychological and Cognitive Disposition*

Alongside factors of efficiency and convenience, the reviewed literature emphasized the importance of mental elements in building trust. Vulnerabilities related to psychological and cognitive disposition were noted in 22 (73.3%) studies. The most common subtheme was anxiety or depression related to disease uncertainty (15/30, 50%), with participants experiencing emotional distress over disease progression or the reliability of remote diagnosis. Cognitive impairment affecting decision-making was reported in 10 (33.3%) papers, encompassing memory decline, reduced attention span, and difficulty comprehending medical information—factors that hindered engagement with telemedicine processes (eg, scheduling appointments, following remote care plans). Low self-esteem or grief-related distress appeared in 6 (20%) studies, where negative self-perceptions (eg, feeling "technologically incompetent") or grief (eg, loss of independence) further reduced willingness to adopt telemedicine.

### *Complex Health Conditions*

Older adults' prevalent physical disabilities and complex health conditions present distinct vulnerabilities within the telemedicine context (19/30, 63.3%), manifesting through 3 interconnected dimensions.

Twelve (40%) papers addressed older adults' chronic multimorbidity, including concurrent conditions such as diabetes, hypertension, and cardiovascular disease, which increased the complexity of remote health monitoring and management. Ten (33.3%) papers reported progressive physical decline, characterized by age-related frailty, reduced mobility, and heightened fall risk—factors that limited independent use of telemedicine devices (eg, wearable sensors and video consultation tools). Seven (23.3%) papers included complications requiring frequent care adjustments, such as acute exacerbations of chronic obstructive pulmonary disease or kidney disease, which necessitated timely in-person clinical intervention and reduced reliance on telemedicine alone. Older patients' heightened awareness of health risks and self-perceived vulnerability drive stringent expectations for health care accuracy, timeliness, and reliability (2/30, 6.7%).

Table 2 presents the co-documentation frequencies between the 4 trust barrier themes and 4 vulnerability dimensions across the 30 included studies. Each cell indicates the number of studies that simultaneously reported both constructs, based on the thematic coding documented in Multimedia Appendix 1. The high co-documentation frequencies across most cells reflect the interconnected nature of these constructs: vulnerabilities rarely map onto a single barrier in isolation, and multiple vulnerability dimensions frequently co-occur with each trust barrier theme within the same study. These patterns are consistent with the narrative findings, which emphasize that the relationships between trust barriers and vulnerabilities are layered and nonlinear rather than reducible to discrete one-to-one pairings.

**Table .** Co-documentation frequency matrix: trust barrier themes×vulnerability dimensions across included studies (N=30)[a].

| Dimension | T1: technophobia and technical difficulties (n=28), n (%) | T2: preference for in-person care (n=26), n (%) | T3: privacy and data concerns (n=23), n (%) | T4: emotional and social impacts (n=21), n (%) |
|---|---|---|---|---|
| V1: low telemedicine literacy (n=29) | 27 (90.0) | 25 (83.3) | 22 (73.3) | 21 (70.0) |
| V2: dependence on social support (n=29) | 27 (90.0) | 25 (83.3) | 22 (73.3) | 20 (66.7) |
| V3: psychological and cognitive disposition (n=22) | 22 (73.3) | 19 (63.3) | 18 (60.0) | 16 (53.3) |
| V4: complex health conditions (n=19) | 18 (60.0) | 16 (53.3) | 14 (46.7) | 13 (43.3) |

[a]Cell values represent the number (and percentage) of included studies simultaneously documenting each trust barrier-vulnerability pair, based on thematic codes in Multimedia Appendix 1. Marginal totals in parentheses indicate how many studies documented each construct independently; co-documentation frequencies should be interpreted in relation to these base rates. No directionality or causal relationship is implied.

## Discussion

### Principal Findings

The most significant finding of this scoping review is the identification of 4 interrelated vulnerability dimensions—technological literacy, health status, psychological or cognitive dispositions, and social support—that collectively underpin trust erosion in older adults' telemedicine use. By synthesizing evidence from previously fragmented studies, this review demonstrates that these vulnerabilities consistently align with 4 clusters of trust barriers: technophobia, privacy concerns, emotional and social impacts, and preference for in-person care. These vulnerabilities operate across 3 interconnected levels: at the individual level, low technological literacy and adverse psychological dispositions restrict autonomy; at the interpersonal level, diminished social support and technology dependence shift erode dignity and emotional validation; and at the systemic level, gaps in platform accessibility, privacy safeguards, and integration with in-person care exacerbate digital exclusion.

Together, these findings offer an integrative and organizing framework that consolidates previously fragmented evidence into a multilevel map of how vulnerability shapes trust dynamics in older adults' telemedicine use. Additionally, the synthesis provides a practical scaffold for targeted intervention design. Corresponding strategies were identified at each level, as detailed in subsequent sections.



XSL•FO
RenderX

### Implications of Telemedicine Trust

Building on this framework, the evidence suggests that older adults' trust in telemedicine is eroded through layered "loss of control"—rooted in interconnected vulnerabilities across all 3 levels [64,72,73].

At the individual level, the loss of bodily control is central. Declining physical functioning increases reliance on technological devices and caregiver assistance, eroding bodily and informational autonomy. Telemedicine tools often feel intrusive (eg, forced disclosure of sensitive information) or impersonal, with standardized services failing to address complex needs. Uncertainty about remote diagnosis further weakens perceived control over health outcomes, directly undermining trust [73,74].

At the interpersonal level, the loss of identity control is pronounced. As older adults withdraw from the labor market, dependence on others for technology use redefines them from "independent" to "cared-for," threatening dignity—particularly among marginalized groups (eg, low educational attainment) [75-77]. Unlike in-person care, virtual interactions reduce emotional validation and social role confirmation, leading to feelings of loneliness, neglect, or diminished attachment [78,79].

At the systemic level, the loss of generational control reflects broader technological shifts. Telemedicine amplifies "digital exclusion" for older adults who are unfamiliar with platforms or cybersecurity, triggering anxiety [53,80,81]. Generational gaps in technological exposure deepen this divide: many individuals prefer in-person care, feeling sidelined in a system that prioritizes digital innovation over familiar health care relationships, further exacerbating their vulnerability [63,82].

In summary, the factors tied to individual autonomy, interpersonal identity, and systemic adaptability demonstrate that trust cannot be addressed through isolated fixes. The multilevel framework provides a cohesive foundation for targeted interventions across all 3 levels. Since this study broadens the age-based definition of "older adults," this framework is better suited as a descriptive map applicable to the broader older adult spectrum rather than a precision model for any single age cohort. In practice, clinicians and policymakers should develop tailored strategies for specific age subgroups rather than treating "older adults" as a homogeneous category.

### Enhancing Inclusive Telemedicine

Translating the multilevel framework into practice, the following recommendations address trust barriers at each level, organized by the strength of evidence.

First, design and training can directly reduce individual barriers to trust. The included studies identified interface complexity and insufficient training as primary drivers of technophobia. To address technological vulnerabilities, telemedicine platforms should prioritize age-friendly interfaces with large fonts, voice-activated commands, and modular operation guides to improve usability [83-85]. Crucially, the included studies also emphasized that on-demand technical support must accompany these tools, as older adults' reluctance to engage often stems not from fixed incapacity but from a fear of making errors without accessible help.

Second, humanizing interactions can strengthen interpersonal connections, which are essential for trust. Multiple studies have identified the loss of the therapeutic relationship and the absence of nonverbal communication as central to emotional distrust in video consultations, pointing to the need for provider training in virtual communication skills. Training providers to use nonverbal cues (eg, nodding and expressive facial gestures) during video consultations adds warmth to virtual visits [86,87]. The evidence also consistently highlights the role of family members and informal caregivers. Involving authorized family members in virtual visits and providing shared record access can reduce recovery anxiety and sustain engagement, thereby reinforcing trust [88,89].

Third, systemic improvements must anchor these efforts, encompassing privacy, cultural adaptability, and accessibility [90,91]. Studies across multiple contexts identified privacy transparency as a foundational trust mechanism. To protect privacy, telemedicine platforms should provide visual guides that clarify the scope of data collection and enable users to customize sharing permissions through intuitive privacy dashboards [91-93]. Sensitive data (eg, mental health records or sexual health information) require double encryption to establish the credibility of data protection. Culturally and linguistically adapted services were also documented as important equity considerations. Multilingual interfaces (eg, Chinese and Spanish) and culturally relevant scheduling (eg, adjustments for Lunar New Year) can accommodate diverse linguistic and cultural needs.

These solutions must be reinforced by broader policy frameworks and practical support systems to achieve scalable impact. The evidence supports subsidizing community technology training and retaining low-tech alternatives (such as telephone-based consultations and printed care instructions) to ensure access for those who cannot or do not wish to use digital channels. Hybrid care models that combine telemedicine with in-person options were endorsed across multiple studies as a pragmatic means of accommodating diverse needs without forcing a binary choice. Such interventions align with the WHO's Decade of Healthy Aging (2021‐2030), which emphasizes digital inclusion as a pillar of equitable health care for older adults, aiming to protect their autonomy in digital health contexts.

### Future Directions

This scoping review mapped the dynamics of trust and vulnerability across individual, interpersonal, and systemic levels, highlighting targeted paths for future research. First, longitudinal studies are necessary to track changes in the dynamics of trust in telemedicine among older adults. Future research can investigate how trust in telemedicine evolves over time for older adults, particularly as they gain experience with digital tools.

Second, effectiveness can be better evaluated, such as training programs or age-friendly redesigns, and should be assessed in terms of their ability to reduce trust barriers and improve



XSL•FO
RenderX

telemedicine adoption. Trust and vulnerability scales for telemedicine, based on empirical tests, have yet to be developed, indicating a need for more scientific measurement tools in future research.

Third, comparative cross-cultural and geographic comparisons remain insufficient. Future research should focus on understudied low- and middle-income countries and rural-urban divides to explore how systemic factors (eg, health care infrastructure) mediate trust-vulnerability links, thereby filling gaps in current regional data.

Fourth, future studies should examine how trust barriers differ across telemedicine modalities. The current evidence suggests that video consultations, wearable monitoring devices, telehealth apps, and electronic surveillance technologies may generate distinct trust concerns and vulnerability patterns. Comparative studies examining modality-specific trust dynamics would enable more targeted design and policy recommendations.

### Limitations

This review has certain limitations. The heterogeneous age thresholds used across included studies represent a substantive methodological limitation that affects the precision of the review's conclusions. While justifications exist for this broader inclusion, the synthesis risks conflating the trust and vulnerability profiles of mid-life adults (50-64 y) with those of older adults (75+ y).

The search strategy constitutes a further methodological limitation. By requiring both "trust" and "vulnerability" as mandatory search terms, the review may have systematically excluded studies that address structurally similar phenomena under alternative conceptual labels. The evidence base synthesized here should, therefore, be understood as representing a theoretically bounded subset of the broader literature on older adults' telemedicine engagement. Future reviews should consider expanding search terms to include "digital exclusion," "digital divide," "autonomy," and "resistance" to capture this adjacent evidence.

In addition, most studies focused on primary medical care, overlooking particularly vulnerable populations, such as individuals with severe disabilities or those in underserved areas, thereby offering only a partial perspective on telemedicine experiences.

The restriction to English-language publications represents not only a practical constraint but also a potential source of cultural and regional bias. Studies conducted in non–English-speaking contexts may report trust barriers and vulnerability profiles that differ substantially from the predominantly Western and East Asian evidence synthesized here. Future reviews should deliberately include non-English literature to achieve more equitable geographic representation.

Another limitation concerns the aggregation of findings across heterogeneous telemedicine modalities. This review synthesized evidence from studies examining wearable devices, video consultations, telemedicine applications, electronic surveillance systems, and government digital health services under a unified thematic framework. While this approach enabled the identification of cross-cutting patterns, it necessarily obscures the trust dynamics specific to each modality.

Finally, the predominance of cross-sectional data restricted insight into the temporal evolution of trust and vulnerability.

### Conclusions

This review highlights that trust in telemedicine among older adults is shaped by a complex interplay of vulnerabilities across individual, interpersonal, and systemic levels. These dynamics show that the erosion of trust is not just a result of technological barriers but also involves bodily autonomy, social identity, and the capacity to adapt to systemic change.

To address these challenges, this study proposes an inclusive telemedicine improvement strategy. "Inclusion" in telemedicine does not always mean "high tech"; rather, meaningful inclusion requires interventions tailored to all 3 levels: intuitive tools and training to empower individuals, humanized interactions to strengthen interpersonal bonds, and systemic policies that balance digital innovation with low-tech alternatives and cultural responsiveness. By centering on such layered solutions, telemedicine can evolve into a more equitable tool that respects older adults' diverse needs, improves health care equity, and enhances their healthcare experience by prioritizing trust, not just technology.

## Acknowledgments

The authors declare the use of generative artificial intelligence (GAI) in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GAI tools under full human supervision: proofreading and editing. The GAI tool used was ChatGPT 5.3 and DeepL. Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes.

## Funding

This work was supported by University of Macau (grant numbers MYRG-GRG2025-00039-FSS and MYRG-GRG2024-00083-FSS).

## Conflicts of Interest

None declared.

Multimedia Appendix 1



XSL•FO
RenderX

Study characteristics.
[DOCX File, 46 KB - i-jmr_v15i1e84818_app1.docx ]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 64 KB - i-jmr_v15i1e84818_app2.pdf ]

## References


1. Gu D, Andreev K, Dupre ME. Major trends in population growth around the world. China CDC Wkly 2021 Jul 9;3(28):604-613. [doi: 10.46234/ccdcw2021.160] [Medline: 34594946]
2. Li W, Guo J, Liu W, Tu J, Tang Q. Effect of older adults willingness on telemedicine usage: an integrated approach based on technology acceptance and decomposed theory of planned behavior model. BMC Geriatr 2024 Sep 17;24(1):765. [doi: 10.1186/s12877-024-05361-y] [Medline: 39289605]
3. Weinstein RS, Lopez AM, Joseph BA, et al. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. Am J Med 2014 Mar;127(3):183-187. [doi: 10.1016/j.amjmed.2013.09.032] [Medline: 24384059]
4. Parimbelli E, Bottalico B, Losiouk E, et al. Trusting telemedicine: a discussion on risks, safety, legal implications and liability of involved stakeholders. Int J Med Inform 2018 Apr;112:90-98. [doi: 10.1016/j.ijmedinf.2018.01.012] [Medline: 29500027]
5. Chang A, Xian X, Liu MT, Zhao X. Health communication through positive and solidarity messages amid the COVID-19 pandemic: automated content analysis of Facebook uses. Int J Environ Res Public Health 2022 May 19;19(10):6159. [doi: 10.3390/ijerph19106159] [Medline: 35627696]
6. Perednia DA, Allen A. Telemedicine technology and clinical applications. JAMA 1995 Feb 8;273(6):483-488. [doi: 10.1001/jama.1995.03520300057037] [Medline: 7837367]
7. Gallardo MO, Dela Torre J, Ebardo R. The role of initial trust in the behavioral intention to use telemedicine among Filipino older adults. Gerontol Geriatr Med 2024;10:23337214231222110. [doi: 10.1177/23337214231222110] [Medline: 38196939]
8. Erfanifa A, Ghazizadeh Hashemi SAH, Abbasi Z, et al. Advantages and disadvantages of telemedicine in medicine: a review study. J Zabol Med Sch 2024;7(1):45-55. [doi: 10.18502/jzms.v7i1.17725]
9. Chen Y, Yuan J, Shi L, et al. Understanding the role of technology anxiety in the adoption of digital health technologies (DHTs) by older adults with chronic diseases in Shanghai: an extension of the Unified Theory of Acceptance and Use of Technology (UTAUT) model. Healthcare (Basel) 2024 Jul 16;12(14):1421. [doi: 10.3390/healthcare12141421] [Medline: 39057564]
10. Arfi WB, Nasr IB, Kondrateva G, Hikkerova L. The role of trust in intention to use the IoT in eHealth: application of the modified UTAUT in a consumer context. Technol Forecast Soc Change 2021 Jun;167:120688. [doi: 10.1016/j.techfore.2021.120688]
11. Liu JYW, Sorwar G, Rahman MS, Hoque MR. The role of trust and habit in the adoption of mHealth by older adults in Hong Kong: a healthcare technology service acceptance (HTSA) model. BMC Geriatr 2023 Feb 4;23(1):73. [doi: 10.1186/s12877-023-03779-4] [Medline: 36737712]
12. Jiao W, Chang A, Ho M, Lu Q, Liu MT, Schulz PJ. Predicting and empowering health for Generation Z by comparing health information seeking and digital health literacy: cross-sectional questionnaire study. J Med Internet Res 2023 Oct 30;25:e47595. [doi: 10.2196/47595] [Medline: 37902832]
13. Hamm JA, Möllering G, Darcy K. Integrating focal vulnerability into trust research. J Trust Res 2024 Jul 2;14(2):237-255. [doi: 10.1080/21515581.2024.2375802]
14. Lewicki RJ, McAllister DJ, Bies RJ. Trust and distrust: new relationships and realities. Acad Manage Rev 1998;23(3):438. [doi: 10.2307/259288]
15. Verhoest K, Latusek D, Six F, et al. Trust and distrust in public governance settings: conceptualising and testing the link in regulatory relations. J Trust Res 2024;14(2):127-156. [doi: 10.1080/21515581.2024.2383918] [Medline: 39502082]
16. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12. [doi: 10.1177/1357633X16674087] [Medline: 29320966]
17. Guo Y, Wen T, Yue S, Zhao X, Huang K. The influence of health information attention and app usage frequency of older adults on persuasive strategies in mHealth education apps. Digit Health 2023;9:20552076231167003. [doi: 10.1177/20552076231167003] [Medline: 37021126]
18. Garg V, Brewer J. Telemedicine security: a systematic review. J Diabetes Sci Technol 2011 May 1;5(3):768-777. [doi: 10.1177/193229681100500331] [Medline: 21722592]
19. Berridge C, Grigorovich A. Algorithmic harms and digital ageism in the use of surveillance technologies in nursing homes. Front Sociol 2022;7:957246. [doi: 10.3389/fsoc.2022.957246] [Medline: 36189442]
20. Bloom G, Berdou E, Standing H, Guo Z, Labrique A. ICTs and the challenge of health system transition in low and middle-income countries. Global Health 2017 Aug 7;13(1):56. [doi: 10.1186/s12992-017-0276-y] [Medline: 28784144]




XSL•FO
RenderX

21. Kaihlanen AM, Virtanen L, Buchert U, et al. Towards digital health equity—a qualitative study of the challenges experienced by vulnerable groups in using digital health services in the COVID-19 era. BMC Health Serv Res 2022 Feb 12;22(1):188. [doi: 10.1186/s12913-022-07584-4] [Medline: 35151302]
22. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172. [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
23. El-Wajeeh M, Galal-Edeen GH, Mokhtar D. Technology acceptance model for mobile health systems. IOSR J Mob Comput Appl 2014;1(1):21-33. [doi: 10.9790/0050-0112133]
24. Bertolazzi A, Quaglia V, Bongelli R. Barriers and facilitators to health technology adoption by older adults with chronic diseases: an integrative systematic review. BMC Public Health 2024 Feb 16;24(1):506. [doi: 10.1186/s12889-024-18036-5] [Medline: 38365698]
25. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556. [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
26. Emerson AJ, Einhorn L, Groover M, Naze G, Baxter GD. Clinical conversations in the management of chronic musculoskeletal pain in vulnerable patient populations: a meta-ethnography. Disabil Rehabil 2023 Oct;45(21):3409-3434. [doi: 10.1080/09638288.2022.2130447] [Medline: 36205554]
27. Rocque R, Leanza Y. A systematic review of patients' experiences in communicating with primary care physicians: intercultural encounters and a balance between vulnerability and integrity. PLoS One 2015;10(10):e0139577. [doi: 10.1371/journal.pone.0139577] [Medline: 26440647]
28. Li X, Huang L, Zhang H, Liang Z. Enabling telemedicine from the system-level perspective: scoping review. J Med Internet Res 2025 Mar 5;27:e65932. [doi: 10.2196/65932] [Medline: 40053725]
29. Western MJ, Smit ES, Gültzow T, et al. Bridging the digital health divide: a narrative review of the causes, implications, and solutions for digital health inequalities. Health Psychol Behav Med 2025;13(1):2493139. [doi: 10.1080/21642850.2025.2493139] [Medline: 40276490]
30. Fothergill L, Hayes N, Latham Y, Hamilton J, Ahmed S, Holland C. Understanding how, for whom and under what circumstances telecare can support independence in community-dwelling older adults: a realist review. BMC Geriatr 2025 Jan 27;25(1):59. [doi: 10.1186/s12877-024-05650-6] [Medline: 39871177]
31. Ge H, Li J, Hu H, Feng T, Wu X. Digital exclusion in older adults: a scoping review. Int J Nurs Stud 2025 Aug;168:105082. [doi: 10.1016/j.ijnurstu.2025.105082] [Medline: 40279791]
32. Seifert A. The digital exclusion of older adults during the COVID-19 pandemic. J Gerontol Soc Work 2020;63(6-7):674-676. [doi: 10.1080/01634372.2020.1764687] [Medline: 32401181]
33. Pham MT, Rajić A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. Res Synth Methods 2014 Dec;5(4):371-385. [doi: 10.1002/jrsm.1123] [Medline: 26052958]
34. Rumrill PD, Fitzgerald SM, Merchant WR. Using scoping literature reviews as a means of understanding and interpreting existing literature. Work 2010 Apr;35(3):399-404. [doi: 10.3233/WOR-2010-0998] [Medline: 20364059]
35. Kastner M, Herrington I, Makarski J, et al. Interventions that have potential to help older adults living with social frailty: a systematic scoping review. BMC Geriatr 2024 Jun 15;24(1):521. [doi: 10.1186/s12877-024-05096-w] [Medline: 38879489]
36. Kastner M, Tricco AC, Soobiah C, et al. What is the most appropriate knowledge synthesis method to conduct a review? Protocol for a scoping review. BMC Med Res Methodol 2012 Aug 3;12:1-10. [doi: 10.1186/1471-2288-12-114] [Medline: 22862833]
37. Tricco AC, Lillie E, Zarin W, et al. A scoping review on the conduct and reporting of scoping reviews. BMC Med Res Methodol 2016 Feb 9;16(1):15. [doi: 10.1186/s12874-016-0116-4] [Medline: 26857112]
38. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143. [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
39. Xian X, Chang A, Xiang YT, Liu MT. Debate and dilemmas regarding generative AI in mental health care: scoping review. Interact J Med Res 2024 Aug 12;13(1):e53672. [doi: 10.2196/53672] [Medline: 39133916]
40. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8(1):45. [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
41. Helou MA, DiazGranados D, Ryan MS, Cyrus JW. Uncertainty in decision making in medicine: a scoping review and thematic analysis of conceptual models. Acad Med 2020 Jan;95(1):157-165. [doi: 10.1097/ACM.0000000000002902] [Medline: 31348062]
42. Hamel C, Michaud A, Thuku M, et al. Defining rapid reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews. J Clin Epidemiol 2021 Jan;129:74-85. [doi: 10.1016/j.jclinepi.2020.09.041] [Medline: 33038541]
43. Tay KT, Tan XH, Tan LHE, et al. A systematic scoping review and thematic analysis of interprofessional mentoring in medicine from 2000 to 2019. J Interprof Care 2021;35(6):927-939. [doi: 10.1080/13561820.2020.1818700] [Medline: 33290115]




XSL•FO
RenderX

44. Finkelstein S, Sharma U, Furlonger B. The inclusive practices of classroom teachers: a scoping review and thematic analysis. Int J Incl Educ 2021 May 12;25(6):735-762. [doi: 10.1080/13603116.2019.1572232]
45. Essex R, Kalocsányiová E, Rumyantseva N, Jameson J. Trust amongst refugees in resettlement settings: a systematic scoping review and thematic analysis of the literature. Int Migration & Integration 2022 Jun;23(2):543-568. [doi: 10.1007/s12134-021-00850-0]
46. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
47. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310] [Medline: 843571]
48. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-269. [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
49. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
50. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339(jul21 1):b2700. [doi: 10.1136/bmj.b2700] [Medline: 19622552]
51. Essén A. The two facets of electronic care surveillance: an exploration of the views of older people who live with monitoring devices. Soc Sci Med 2008 Jul;67(1):128-136. [doi: 10.1016/j.socscimed.2008.03.005] [Medline: 18396367]
52. Fitzsimmons DA, Thompson J, Bentley CL, Mountain GA. Comparison of patient perceptions of telehealth-supported and specialist nursing interventions for early stage COPD: a qualitative study. BMC Health Serv Res 2016 Aug 22;16(1):420. [doi: 10.1186/s12913-016-1623-z] [Medline: 27549751]
53. Kruse C, Fohn J, Wilson N, Nunez Patlan E, Zipp S, Mileski M. Utilization barriers and medical outcomes commensurate with the use of telehealth among older adults: systematic review. JMIR Med Inform 2020 Aug 12;8(8):e20359. [doi: 10.2196/20359] [Medline: 32784177]
54. Ladin K, Porteny T, Perugini JM, et al. Perceptions of telehealth vs in-person visits among older adults with advanced kidney disease, care partners, and clinicians. JAMA Netw Open 2021 Dec 1;4(12):e2137193. [doi: 10.1001/jamanetworkopen.2021.37193] [Medline: 34870680]
55. Schorr EN, Gepner AD, Dolansky MA, et al. Harnessing mobile health technology for secondary cardiovascular disease prevention in older adults: a scientific statement from the American Heart Association. Circ Cardiovasc Qual Outcomes 2021 May;14(5):e000103. [doi: 10.1161/HCQ.0000000000000103] [Medline: 33793309]
56. Chen N, Liu P. Assessing elderly user preference for telehealth solutions in China: exploratory quantitative study. JMIR Mhealth Uhealth 2022 Jan 12;10(1):e27272. [doi: 10.2196/27272] [Medline: 35019852]
57. Hunter I, Lockhart C, Rao V, Tootell B, Wong S. Enabling rural telehealth for older adults in underserved rural communities: focus group study. JMIR Form Res 2022 Nov 4;6(11):e35864. [doi: 10.2196/35864] [Medline: 36331533]
58. Singh N, Misra R, Singh S, Rana NP, Khorana S. Assessing the factors that influence the adoption of healthcare wearables by the older population using an extended PMT model. Technol Soc 2022 Nov;71:102126. [doi: 10.1016/j.techsoc.2022.102126]
59. Korkmaz Yaylagul N, Kirisik H, Bernardo J, et al. Trends in telecare use among community-dwelling older adults: a scoping review. Int J Environ Res Public Health 2022 Dec 12;19(24):16672. [doi: 10.3390/ijerph192416672] [Medline: 36554553]
60. Orzechowski M, Skuban-Eiseler T, Ajlani A, Lindemann U, Klenk J, Steger F. User perspectives of geriatric German patients on smart sensor technology in healthcare. Sensors (Basel) 2023 Nov 11;23(22):9124. [doi: 10.3390/s23229124] [Medline: 38005512]
61. Ali SM, Gambin A, Chadwick H, Dixon WG, Crawford A, Van der Veer SN. Strategies to optimise the health equity impact of digital pain self-reporting tools: a series of multi-stakeholder focus groups. Int J Equity Health 2024 Nov 11;23(1):233. [doi: 10.1186/s12939-024-02299-w] [Medline: 39529006]
62. Chu G, Silva C, Adams K, et al. Exploring the factors affecting home dialysis patients' participation in telehealth-assisted home visits: A mixed-methods study. J Ren Care 2024 Jun;50(2):128-137. [doi: 10.1111/jorc.12475] [Medline: 37434485]
63. Fang Z, Liu Y, Peng B. Empowering older adults: bridging the digital divide in online health information seeking. Humanit Soc Sci Commun 2024;11(1). [doi: 10.1057/s41599-024-04312-7]
64. Niu S, Hong W, Ma Y. How expectations and trust in telemedicine contribute to older adults' sense of control: an empirical study. Healthcare (Basel) 2024 Aug 23;12(17):1685. [doi: 10.3390/healthcare12171685] [Medline: 39273710]
65. Turcotte S, Bouchard C, Rousseau J, et al. Factors influencing older adults' participation in telehealth interventions for primary prevention and health promotion: a rapid review. Australas J Ageing 2024 Mar;43(1):11-30. [doi: 10.1111/ajag.13244] [Medline: 38014903]
66. Windle A, Marshall A, de la Perrelle L, et al. Factors that influence the implementation of innovation in aged care: a scoping review. JBI Evid Implement 2024 Feb 1;22(1):61-80. [doi: 10.1097/XEB.0000000000000407] [Medline: 38153118]




XSL•FO
RenderX

67. Husain L, Greenhalgh T. Examining Intersectionality and barriers to the uptake of video consultations among older adults from disadvantaged backgrounds with limited English proficiency: qualitative narrative interview study. J Med Internet Res 2025 Jan 6;27:e65690. [doi: 10.2196/65690] [Medline: 39761566]
68. Tan SH, Yap YY, Tan SK, Wong CK. Determinants of telehealth adoption among older adults: cross-sectional survey Study. JMIR Aging 2025 Mar 24;8:e60936. [doi: 10.2196/60936] [Medline: 40126531]
69. Zhu Y, Xiao QE, Ao MC, Zhao X. How eHealth use and cancer information-seeking influence older adults' acceptance of genetic testing: mediating roles of PIGI and cancer worry. Digit Health 2025;11:20552076251317658. [doi: 10.1177/20552076251317658] [Medline: 39896267]
70. Adams E, Donaghy E, Sanders C, et al. Digital interventions for older people experiencing homelessness: systematic scoping review. J Med Internet Res 2025 Feb 21;27:e63898. [doi: 10.2196/63898] [Medline: 39984162]
71. Zhang C, Mohamad E, Azlan AA, Wu A, Ma Y, Qi Y. Social media and eHealth literacy among older adults: systematic literature review. J Med Internet Res 2025 Mar 26;27:e66058. [doi: 10.2196/66058] [Medline: 40138684]
72. Cimperman M, Brenčič MM, Trkman P, de Leonni Stanonik M. Older adults' perceptions of home telehealth services. Telemed J E Health 2013 Oct;19(10):786-790. [doi: 10.1089/tmj.2012.0272] [Medline: 23931702]
73. Charness N, Boissy P, Demiris G, Krupinski EA, Lai AM, Lopez AM. How human factors can influence the elderly in the use of telemedicine. Telemed J E Health 2010 Oct;16(8):860-866. [doi: 10.1089/tmj.2010.9948] [Medline: 20958209]
74. Hensel BK, Demiris G, Courtney KL. Defining obtrusiveness in home telehealth technologies: a conceptual framework. J Am Med Inform Assoc 2006;13(4):428-431. [doi: 10.1197/jamia.M2026] [Medline: 16622166]
75. Clancy A, Simonsen N, Lind J, Liveng A, Johannessen A. The meaning of dignity for older adults: a meta-synthesis. Nurs Ethics 2021 Sep;28(6):878-894. [doi: 10.1177/0969733020928134] [Medline: 32613895]
76. Draper P. Dignity, well-being and identity: a qualitative study of older people's interpretations of how healthcare workers speak to them. Nurs Older People 2021 Feb 2;33(1):21-27. [doi: 10.7748/nop.2020.e1261] [Medline: 32935501]
77. Huang L, Song X, Liu MT, Chang WY, Shi GJ. Placebo effect in food marketing: young consumers' perception of food healthiness and tastiness in reduced-sugar labeling. Asia Pac J Mark Logist 2024 Jul 10;36(7):1595-1615. [doi: 10.1108/APJML-07-2023-0700]
78. Pfeil U, Svangstu K, Ang CS, Zaphiris P. Social roles in an online support community for older people. Int J Hum Comput Interact 2011 Feb 23;27(4):323-347. [doi: 10.1080/10447318.2011.540490]
79. Tsang VHL, Tse DCK, Chu L, Fung HH, Mai C, Zhang H. The mediating role of loneliness on relations between face-to-face and virtual interactions and psychological well-being across age: a 21-day diary study. Int J Behav Dev 2022 Nov;46(6):500-509. [doi: 10.1177/01650254221132775]
80. Hider S, Muller S, Gray L, et al. Digital exclusion as a potential cause of inequalities in access to care: a survey in people with inflammatory rheumatic diseases. Rheumatol Adv Pract 2023;7(1):rkac109. [doi: 10.1093/rap/rkac109] [Medline: 36632437]
81. Liu PL, Chang A, Liu MT, et al. Effect of information encounter on concerns over healthy eating- mediated through body comparison and moderated by body mass index or body satisfaction. BMC Public Health 2023 Feb 6;23(1):254. [doi: 10.1186/s12889-023-15069-0] [Medline: 36747209]
82. Lopez de Coca T, Moreno L, Alacreu M, Sebastian-Morello M. Bridging the generational digital divide in the healthcare environment. J Pers Med 2022 Jul 26;12(8):1214. [doi: 10.3390/jpm12081214] [Medline: 35893307]
83. Wong Y, Neo XS. Smart cities for aging populations: future trends in age-friendly public health policies. J Foresight Health Gov 2025;2(1):11-20. [doi: 10.61838/jfph.2.1.2]
84. Tha SH, Hough K, Bui N, Dulaney S, Perissinotto C. The use of age-friendly technology in the care of older adults: two implementation case studies. J Appl Gerontol 2026 Mar;45(3):515-526. [doi: 10.1177/07334648251343513] [Medline: 40552801]
85. Holliday AM, Hashmi AZ, Okoli-Umeweni AO, et al. American Geriatrics Society position statement: telehealth policy for older adults. J Am Geriatr Soc 2025 Dec;73(12):3646-3654. [doi: 10.1111/jgs.70004] [Medline: 40650623]
86. Rahmanti AR, Yang HC, Huang CW, et al. Validating nonverbal cues for assessing physician empathy in telemedicine: a Delphi study. Med Educ Online 2025 Dec;30(1):2497328. [doi: 10.1080/10872981.2025.2497328] [Medline: 40338675]
87. Guan X, Porter MC, Omodt PG. Patient experience diagnosis: using telemed simulation to assess health care provider verbal and nonverbal communication issues to prescribe potential interventions. Patient Exp J 2024;11(1):53-64. [doi: 10.35680/2372-0247.1832]
88. White J, Byles J, Walley T. The qualitative experience of telehealth access and clinical encounters in Australian healthcare during COVID-19: implications for policy. Health Res Policy Syst 2022 Jan 15;20(1):9. [doi: 10.1186/s12961-021-00812-z] [Medline: 35033107]
89. Lu Q, Schulz PJ, Chang A. Medication safety perceptions in China: media exposure, healthcare experiences, and trusted information sources. Patient Educ Couns 2024 Jun;123:108209. [doi: 10.1016/j.pec.2024.108209] [Medline: 38367304]
90. Song X, Liu MT, McCartney GJ, Xian X, Chang A. Revisiting privacy in interactive marketing: perspectives from Asian consumers. J Res Interact Mark 2025;19(8):1223-1243. [doi: 10.1108/JRIM-06-2024-0284]




XSL•FO
RenderX

91. Pool J, Akhlaghpour S, Fatehi F, Gray LC. Data privacy concerns and use of telehealth in the aged care context: an integrative review and research agenda. Int J Med Inform 2022 Apr;160:104707. [doi: 10.1016/j.ijmedinf.2022.104707] [Medline: 35131698]
92. Mishkin AD, Zabinski JS, Holt G, Appelbaum PS. Ensuring privacy in telemedicine: ethical and clinical challenges. J Telemed Telecare 2023 Apr;29(3):217-221. [doi: 10.1177/1357633X221134952] [Medline: 36349356]
93. Chang A, Schulz PJ, Jiao W, Liu MT. Obesity-related communication in digital Chinese news from Mainland China, Hong Kong, and Taiwan: automated content analysis. JMIR Public Health Surveill 2021 Nov 23;7(11):e26660. [doi: 10.2196/26660] [Medline: 34817383]

## Abbreviations

**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
**RQ:** research question
**WHO:** World Health Organization

*Edited by M Balcarras; submitted 25.Sep.2025; peer-reviewed by MI Rahman, TAK Manne; revised version received 01.Jun.2026; accepted 01.Jun.2026; published 14.Jul.2026.*

*Please cite as:*
*Wang W, Ma X, Liu MTC, Chang A*
*Trust Barriers and Vulnerabilities in Older Adults' Telemedicine Adoption: Scoping Review*
*Interact J Med Res 2026;15:e84818*
*URL: https://www.i-jmr.org/2026/1/e84818*
*doi:10.2196/84818*

© Wanqing Wang, Xin Ma, Matthew Ting Chi Liu, Angela Chang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 14.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Digital Health Interventions Incorporating Behavior Change Techniques and Gamification for Bladder Health in Adults Aged 50 Years and Older: Rapid Review

Oscar Aguila-Gimeno[1,2,3], MSc; Norina Gasteiger[4], PhD; Emma Stanmore[4], PhD; Marius Brazaitis[5], PhD; Rima Solianik[5], PhD; Erika Karkauskiene[5], MD; Laura Jarutienė[5], MD; Júlia Romeu-Busquets[1], MSc; Javier Jerez-Roig[1,2,5], PhD

[1]Faculty of Health Sciences and Welfare, Research Group on Methodology, Methods, Models and Outcomes of Health and Social Sciences ($M_3O$), Centre for Health and Social Care Research (CESS). University of Vic-Central University of Catalonia (UVic-UCC), Vic, Spain

[2]Institute for Research and Innovation in Life Sciences and Health in Central Catalonia (IRIS-CC), Vic, Spain

[3]Hospital Universitari de Santa Maria, Lleida, Spain

[4]School of Health Sciences, The University of Manchester, Oxford Rd, Manchester, Manchester, United Kingdom

[5]Institute of Sport Science and Innovations, Lithuanian Sports University, Kaunas, Lithuania, Lithuania

**Corresponding Author:**
Emma Stanmore, PhD
School of Health Sciences, The University of Manchester, Oxford Rd, Manchester, Manchester, United Kingdom

## Abstract

**Background:** Urinary incontinence (UI) is the most prevalent pelvic floor dysfunction, with incidence increasing with age. Numerous studies have demonstrated the effectiveness of pelvic floor exercises in improving UI. However, access to pelvic floor treatment remains limited due to lengthy waiting lists and poor adherence to prescribed exercises. Digital solutions incorporating behavior change techniques (BCTs), including gamification elements, may support self-management of bladder health for individuals aged 50 years and older who may face challenges in accessing conventional treatments.

**Objective:** This study aimed to identify mobile apps and websites integrating BCTs, including gamification elements, to facilitate bladder health self-management among adults aged 50 years and older.

**Methods:** The search was initiated in July 2024 and updated in January 2025 across three sources: (1) mobile app stores (Google Play Store and Apple App Store available in Spain, Lithuania, and the United Kingdom) using the keywords “pelvic floor,” “urinary incontinence,” and “bladder”; (2) websites; and (3) academic databases (PubMed, Scopus, CINAHL, and Google Scholar) for journal articles, book chapters, and conference papers published in English, Spanish, or Lithuanian. Inclusion criteria required that solutions be independently usable without health care supervision, incorporate at least one BCT (eg, training, education), be evidence-informed (ie, include participants aged ≥50 years in their design or piloting phase or explicitly target this demographic), and be available in Spain, the United Kingdom, or Lithuania. The Mobile App Rating Scale (MARS) was used to assess app quality, and the taxonomy of 93 BCTs was used for BCT classification. All review and data extraction processes were conducted in duplicate.

**Results:** Twenty-one studies met the inclusion criteria, identifying 8 eligible mobile apps and 1 website. Among the apps, only 2 were available on either Google Play or the Apple App Store. No apps were identified in Lithuania, whereas 1 app was found in Spain and 3 apps in the United Kingdom. Of these, 1 UK app was accessible on both Google Play and the Apple App Store, whereas the others were limited to a single platform. BCT extraction showed that the apps included between 9 and 16 BCTs (mean 12, SD 3.27). Regarding quality, all assessed apps obtained MARS scores ranging from 3 to 4 out of 5. Website searches did not identify any scientifically validated platforms across the 3 countries, except for 1 website cited in a scientific publication. UI reduction on the International Consultation on Incontinence Questionnaire ranged from –3.9 to –2.1 points, while perceived improvement reached 91.7% in the Tät app.

**Conclusions:** Evidence-based digital interventions for individuals aged 50 years and older remain limited. Existing apps suggest potential benefits in UI reduction and quality of life improvement; however, further research and development are needed to enhance accessibility and efficacy.

**Trial Registration:** PROSPERO CRD42024597624; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024597624

***(Interact J Med Res 2026;15:e76867)*** doi:10.2196/76867



XSL•FO
RenderX

**KEYWORDS**

urology; gynaecology; urinary incontinence; pelvic floor exercises; pelvic floor muscle training; behavior change; digital health; healthy aging; gamification

## Introduction

The incidence of pelvic floor dysfunction is increasing, especially among older adults. Urinary incontinence (UI) is the most common dysfunction of the pelvic floor, and the incidence increases with age. UI is the involuntary loss of urine, affecting up to 36 million people in Europe [1-3].

The total direct cost in Germany, Italy, Spain, Sweden, and the United Kingdom has been estimated at €5 billion per year (€1=US $1.14 as of July 12, 2026), and this burden will increase with the projected 25% increase in UI prevalence by 2030 [2]. The mean annual per capita cost of its treatment is €300 - €600 per country [2].

The most frequent types of UI are stress urinary incontinence, which is characterized by the involuntary leakage of urine during activities that increase intraabdominal pressure, such as coughing or sneezing; urgency urinary incontinence, defined by a sudden and compelling need to urinate, usually accompanied by leakage before reaching the toilet; and mixed urinary incontinence, which combines features of both previous types [4].

Several studies have explored the effectiveness of pelvic floor exercises to improve UI and reduce the involuntary loss of urine [5,6]. A combination of behavioral therapies consisting of pelvic floor muscle training, bladder training, and practical tips is beneficial in the management of UI [5]. Also, several studies have indicated that low levels of physical activity (PA) and prolonged patterns of sedentary behavior represent risk factors for UI [7,8], with PA being defined by the World Health Organization as any bodily movement produced by skeletal muscles that requires energy expenditure [9]. Some individuals with UI cannot access pelvic floor treatment because there are barriers to uptake, including waiting lists that limit access to treatment, poor adherence to exercises [1,10], the price of the treatment, or embarrassment.

Ultimately, pelvic floor dysfunctions can reduce the quality of life of older people, with research showing that several interrelated factors can challenge the independence of older adults: primarily functional and cognitive impairment, chronic diseases, a diminishing social network, and a low level of PA [11].

Gamification should not be understood merely as the incorporation of game elements into nongame contexts, but rather as a process aimed at generating meaningful psychological experiences that enhance the perceived value for the user. This emphasizes that motivational affordances (such as points, levels, or badges) function in nongame contexts as stimuli designed to activate psychological needs and emotional states, with the purpose of influencing behavior and strengthening emotional engagement [12].

In contrast to gamification, video games constitute complete, self-contained interactive environments designed primarily for entertainment. While video games involve full game narratives, mechanics, and immersive worlds, gamification selectively integrates specific game design elements into health or behavioral interventions without necessarily creating a full game experience. This conceptual distinction is essential, as gamified interventions and video-game–based interventions may rely on different psychological mechanisms, levels of immersion, and theoretical foundations [13,14].

Grounded in motivational psychology, gamification fosters constructs such as self-efficacy, engagement, and habit formation, particularly through the promotion of autonomy, competence, and social relatedness [13]. These factors are directly linked to self-determination theory, which highlights the importance of satisfying basic psychological needs in order to sustain intrinsic motivation [13].

Furthermore, gamification is conceived as an attempt to increase the likelihood of the emergence of experiences of flow, challenge, or immersion [14], rather than as a fixed set of game mechanics. In this way, psychology becomes the central axis explaining how design elements are transformed into emotional and motivational experiences, which ultimately determine the effectiveness of gamification in creating value for the user [12].

In recent years, emerging evidence suggests that gamification may serve as a valuable tool in addressing health-related challenges among patients [15]. According to the meta-analysis conducted by Mazeas et al [15], gamified interventions are not only effective in modifying behavior, but also demonstrate greater efficacy compared to other behavioral approaches. Moreover, these interventions show promise in supporting the long-term maintenance of behavioral changes.

Another review highlights that supervised exercise programs incorporating game elements appear to increase PA levels more effectively than conventional exercise routines [16].

Behavior change techniques (BCTs) are defined as observable, replicable, and irreducible components of an intervention designed to modify behavior [17]. They represent the smallest active ingredients that can be applied individually or in combination to influence behavioral outcomes. BCTs, such as feedback or monitoring, can enhance the maintenance of behavior over time through instant feedback and improve performance in eHealth programs [18]. Behavioral treatment for pelvic floor dysfunctions can be an effective tool either on its own or in combination with medication to improve UI, urinary urgency, nocturia, and other conditions [19].

In the context of mobile technologies and web-based interventions targeting bladder and pelvic floor health, the application of the BCT taxonomy alongside gamification enables the identification of active ingredients that promote adherence, engagement, and sustained behavior change in adults and older populations.



XSL•FO
RenderX

Given the growing number of pelvic floor apps available online, this review focused on evidence-based programs to assess their actual effectiveness. A 2019 systematic review identified 840 pelvic floor–related apps, of which approximately 78% were not developed by healthcare professionals. This raises concerns regarding the usability and clinical efficacy of apps that lack a foundation in scientific evidence [20].

Rapid reviews have been recognized by the Cochrane Rapid Reviews Methods Group and the World Health Organization as an appropriate approach when timely evidence is needed to inform intervention development in fast-evolving fields [21,22]. In the context of digital health for bladder and pelvic floor self-management, technologies evolve quickly and new apps emerge each year, making a rapid review particularly suitable for providing up-to-date, actionable insights. To ensure methodological rigor despite the accelerated timeline, we followed established rapid review guidance, including the use of predefined eligibility criteria, structured search strategies, and transparent reporting of methodological shortcuts [21]

Technology can offer a solution to some of these challenges, such as UI. The willingness of older people to use technology depends on their perceived need. Initially, older people may believe that technology can be useful for others, but not for them. However, when they try it, they tend to want to continue using it [23]. Due to the increasing prevalence of pelvic floor disorders and the growing importance of technology in health and daily life, it is important to examine mobile technologies and websites designed for self-management among these populations. Therefore, the specific objective of this work was to synthesize the current literature and available apps in Spain, Lithuania, and the United Kingdom for the promotion of bladder health and/or the management of UI among adults aged 50 years and older.

## Methods

### Overview

This review is the first part of the KOKU Bladder project (ClinicalTrials.gov, NCT06583733), an international collaboration involving the University of Vic (Spain), Lithuanian Sports University (Kaunas, Lithuania), and the University of Manchester (United Kingdom). The review included mobile apps, websites, and academic databases from each participating country. Where possible, we adhered to the recommended 7-step method for conducting reviews of mobile health apps [24]. We used a similar approach to combining reviews of the evidence, smartphone apps, and websites as that used in a review conducted by McGarrigle et al [25]. The protocol was registered on October 4, 2024, in PROSPERO (reference CRD42024597624). The Population, Intervention, Comparison, and Outcome (PICO) framework was used to guide the formulation of the inclusion criteria. Pilot searches were conducted in July 2024 to identify appropriate search keywords, followed by the final search between October 2024 and January 2025.

The rapid review was conducted to preliminarily explore the available literature to lay the foundation for the project and guide the initial stages of work. To achieve this, a targeted search methodology was applied, examining a smaller number of databases compared to a traditional systematic review, which allowed for a faster process. In addition to databases, websites and app stores were also reviewed. The total duration of the study was four months, a shorter period than that required for more exhaustive reviews, but sufficient to obtain an overview of the existing evidence and support the project's initial decision-making.

### Eligibility Criteria

The following criteria were used to select relevant digital solutions (apps or websites) across the three searches (Textbox 1).

**Textbox 1.** Inclusion and exclusion criteria.

**Inclusion criteria**:

- Apps or websites that can be used by people independently, without supervision by a healthcare professional.
- Digital technologies (apps and websites) regardless of cost (eg, to download or to access in-app functions).
- Interventions including at least one behavior change technique (eg, training, education, monitoring, and prompting).
- Evidence-informed: resources appropriate for use with older populations (ie, they have included participants aged >50 years in the design or piloting phase or target this group). They do not have to exclusively include those aged >50 years, but rather be part of the study group.
- Available in Spain, the United Kingdom, or Lithuania and in English, Spanish, or Lithuanian.

**Exclusion criteria**:

- Apps or websites designed for pregnant or postpartum individuals, those aged <50 years, or health and social care professionals.
- Apps or websites only including pelvic floor muscle techniques and/or training.
- Apps or websites that require invasive devices (eg, intravaginal devices).
- Intended only for health care professionals.
- For the academic database searches, we excluded review studies, case reports, letters to the editor, editorials, conference abstracts, and personal opinions and/or commentaries.



XSL•FO
RenderX

### Search and Screening

The entire review process was conducted by at least 2 researchers in each participating country. The search and retrieval of scientific papers were carried out by the Spanish research team (OA-G and JJ-R) in October 2024. All retrieved papers were subsequently uploaded to the Rayyan (Rayyan Systems Inc.) bibliographic management platform for review by the research team [26].

At the outset of the review process, duplicate records were identified and removed using Rayyan's automated detection system. Following this, the titles and abstracts of all papers were screened by the research team, ensuring that each paper was reviewed by a minimum of 2 independent reviewers. The selected papers were then downloaded, thoroughly read, and subjected to screening in duplicate. The initial review was conducted by 1 researcher from Spain (OA-G) and 1 from the United Kingdom (ES). A secondary review was carried out by the Lithuanian team (RS, EK, and LJ), comprising three researchers. Upon completion of this phase, a second Spanish researcher examined and resolved any conflicts (JJ-R), leading to the final selection of papers. The extracted data were subsequently organized into tables for analysis.

For the search of mobile apps and websites, at least 2 researchers conducted the process in each country: OA-G, JJ-R, and JRB in Spain; ES and NG in the United Kingdom; and RS, EK, and LJ in Lithuania. The search results were compiled and recorded in tables. Subsequently, OA-G and NG reviewed the findings, while JJ-R resolved any discrepancies or uncertainties.

### App Market

A systematic search was conducted in the Google Play Store and Apple App Store in Spain, Lithuania, and the United Kingdom. Searches were performed on Android (Samsung Galaxy Note 20 Ultra, Android 11; Samsung Galaxy S24, Android 14 OS) and iPhone (iPhone 12, iOS 16.1.1). The search keywords included "pelvic floor," "urinary incontinence," and "bladder." Equivalent terms in Spanish and Lithuanian were used for searches within the respective app markets.

A peer review process was implemented for the app search to ensure consistency and accuracy. The final results from all 3 countries were compiled and compared in a comprehensive data table.

### Websites

We searched for websites featuring interventions aimed at promoting behavior change related to bladder health (eg, UI) with supporting scientific evidence. The search was conducted in 3 countries and 3 languages, with each country searching in its native language. The search strategy was in English included the terms UI exercises, pelvic floor exercises, pelvic floor muscle training, and bladder exercises. In Spanish: ejercicios para la incontinencia urinaria, ejercicios de suelo pélvico, entrenamiento muscular del suelo pélvico, and ejercicios de vejiga. In Lithuanian: šlapimo nelaikymo pratimai, dubens dugno pratimai, dubens dugno raumenų treniravimas, and šlapimo pūslės pratimai.

The search was conducted using Google. We systematically reviewed the first 20 search results from each country for each keyword. This cutoff was based on user behavior research indicating that most users only explore the first few links in search engine results [27]. Rew et al [28] reported that the number of websites reviewed per keyword varies from 10 to 200, with the first 25 websites being a commonly used number.

All the websites identified were reviewed to determine whether they included scientific papers supporting the effectiveness of the site in reducing UI.

### Academic Databases

A systematic search was conducted to identify scientific evidence supporting the included mobile apps and websites. The search was performed in the following academic databases: PubMed, Scopus, CINAHL, and Google Scholar.

Eligible sources included journal papers, book chapters, and full-length conference papers published in English, Spanish, or Lithuanian.

The search strategy in PubMed was (("pelvic floor") OR (urinary incontinence) OR (bladder)) AND ((app*) OR (mHealth) OR (eHealth) OR (game*)), in SCOPUS we limited to paper and aged, and the search strategy was: (("pelvic floor") OR ("urinary incontinence")) ((app*) OR (mhealth) OR (ehealth) OR (technology)). In CINAHL we limited by source type (academic journal) and age (aged 80 years and older and middle-aged: 45 - 64 years years), and the search strategy was (("pelvic floor") OR urinary incontinence OR bladder) AND (app* OR mHealth OR eHealth OR game*)

### Evaluation, Data Extraction, and Synthesis

#### *Evaluation: Mobile App Rating Scale*

The quality of the apps was evaluated using the validated Mobile App Rating Scale (MARS) [29] by 2 researchers in the United Kingdom (ES and NG) and in Spain (OA-G and JR). The criteria assess quality by considering engagement, functionality, esthetics, information quality, and subjective quality, leading to the development of 23 subcategories from which the 23 individual MARS [29] items are derived. Each item is rated on a 5-point scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent), with descriptors provided for each rating anchor. In cases where an item was not applicable, a "not applicable" option was included. In this review, we used the first 19 subcategories and excluded the last category, "subjective quality."

MARS [29] was used in duplicate, and we calculated the mean scores for the final results for each app.

#### *Evaluation: BCTs*

To analyze the BCTs in the apps, we used the taxonomy of BCTs developed by Michie et al [30]. The taxonomy includes a list of 93 BCTs, each evaluated with a score of "0" (absent) or "1" (present). Scores were then summed. The total score for each app ranged from 0 to 93, with higher scores reflecting a greater implementation of BCTs [30].



XSL•FO
RenderX

BCTs were also determined in duplicate, in the United Kingdom (ES and NG) and in Spain (OA-G and JR). Consensus was reached for any apps in which there was disagreement between the researchers.

### *Evaluation: Databases*

During the review process, data from the included studies were extracted systematically into predefined tables. For each intervention study, we recorded key sample characteristics such as mean age, number of participants, educational level, and type of UI, as well as the assessment tools used and the main conclusions reported. This approach allowed us to compare study populations and to examine differences across the validated questionnaires used.

For the included studies, we collected demographic and clinical variables (sample size, age, education level, percentage of women, and UI type) to contextualize the populations targeted by each intervention. These characteristics are essential for assessing the applicability of digital tools to older adults, who often present heterogeneous clinical profiles and varying levels of digital literacy. In addition, we extracted the validated questionnaires used in each study (eg, International Consultation on Incontinence Questionnaire - Short Form (ICIQ-SF), International Consultation on Incontinence Questionnaire - Overactive Bladder (ICIQ-OAB), Patient Global Impression of Improvement (PGI-I), and Pelvic Muscle Physical Therapy (PMPT)), as these instruments represent the most widely accepted clinical and patient-reported outcomes in UI research.

Once the final set of eligible papers was identified, we conducted a structured data extraction process. All studies were read in full, and the relevant information was systematically recorded in predefined tables. For each intervention study, we extracted demographic and clinical characteristics of the sample, including mean age, number of participants, education level, and type of UI, as well as details on study design and the assessment instruments used.

Once the final set of eligible papers was confirmed, we refined the extraction through a structured, full-text review of each study. In parallel, information on the identified mobile apps was collected by analyzing both the app content and the associated scientific publications. App-related data included platform availability, version number, cost, country of origin, and study setting. BCTs were identified using the Behavior Change Technique Taxonomy, app quality was assessed using the MARS [29], and gamification elements were identified through direct inspection of the apps, noting features such as levels, progress tracking, rewards, or visual feedback. All extracted information was organized to ensure consistency and facilitate comparison across studies and apps.

The results were collected in 3 tables. Table 1 contains data related to the apps, the country in which they were found, app store, version number, developer name, price, study setting, number of BCTs, and MARS [29] outcome (Multimedia Appendix 1). Table 2 includes the general information for the selected papers: app name, sample size, mean age, educational level, percentage of women, study design, and percentage of IU (Multimedia Appendix 2). Finally, Table 3 provides the results of the papers and the different questionnaires (Multimedia Appendix 3).

**Table .** Technological and qualitative information of the apps.

| Name of app | Country | App market | Version number | Developer name | Cost | Study setting | BCTs[a] | MARS[b] |
|---|---|---|---|---|---|---|---|---|
| BladderBoss | United Kingdom | Apple App Store | 1.8 | Amara Therapeutics | First week free, after monthly €24.99 | Ireland | 16 | Good 4 (0.20) |
| Tät app | Spain | Google Play | 5.2.7 | eContinence AB | First week free, after €3.19 per month | Sweden | 9 | Acceptable 3.39 (0.02) |
| BladderBoss | United Kingdom | Google Play | 1.0.443 | Amara Therapeutics | First week free, after monthly €24.99 | Ireland | 17 | Acceptable 3.35 (0.05) |
| URinControl | Spain | Google Play | 1.2.5 | Curavista | Free | Sweden and Netherlands | 10 | Acceptable 3.33 (0.07) |
| URinControl | United Kingdom | Google Play | 1.2.5 | Curavista | Free | Sweden and Netherlands | 9 | Acceptable 3.33 (0.07) |
| URinControl4all | United Kingdom | Apple Store | 1.0.0 | Curavista | Free | Netherlands | 11 | Acceptable 3.0 (0.20) |

[a]BCT: behavior change technique.

[b]MARS: Mobile App Rating Scale.



XSL•FO
RenderX

**Table .** General information of the papers.

| Reference | Name of app or webpage | Sample size (control/ intervention group), % (n/N) | Mean age (range or SD) | Education level (%, university or higher) | Women (%, n/N) | Study design | IU[a] types (%, n/N) |
|---|---|---|---|---|---|---|---|
| [31] | Tät II | 123 (60/63) | 58.3 (31-77) | Majority | 100 (123/123) | RCT[b] | MUI:[c] 72 (89/123) and UUI:[d] 22.80 (28/123) |
| [32] | Tät II | 122 (60/62) | 58.3 (SD 9.6) | NR[e] | 100 (122/122) | RCT (cost-utility analysis) | NR |
| [33] | Tät II | 123 (60/63) | 58.6 (49.5 - 67.7) | 61.8 (76/123) | 100 (123/123) | RCT | MUI: 73.5 (90/123) and IUU: 26.5 (33/123) |
| [34] | Tät II | 123 (61/62) | (27-72) Responders at 2 years (mean age 44.2±10.3), and non-responders at 2 years (mean age 46.4, SD 7.8) | NR | 100 (123/123) | Prospective cohort RCT | SUI:[f] 100 (123/123) |
| [35] | URinControl | 262 (131/131) | 53 (20-86) | NR | 100 (262/262) | RCT | MUI: 50 (131/262) |
| [36] | URinControl | 9 | (32-68) | NR | 100 (9/9) | Qualitative | SUI: 11 (1/9), UUI: 33 (3/9), and MUI: 56 (5/9) |
| [37] | URinControl | 17 | (35-78) | NR | 100 (17/17) | Mixed methods | NR |
| [38] | URinControl | 262 (131/131) | (mean 52.2, SD 11.6) | 52.7 (138/262) | 100 (262/262) | Prediction model based RCT | SUI: 68.7 (180/262) and UUI: 31.3 (82/262) |
| [39] | URinControl | 262 (131/131) | 54 (23 - 86) | NR | 100 (262/262) | RCT | SUI: 34 (89/262) |
| [40] | Tät | 2672 (0/2672) | 43.7 | 65.3 (1745/2672) | 100 (2672/2672) | Cohorts study | SUI: 53.1 (1419/2672), UUI: 12.1 (232/2672), and MUI: 30.9 (1021/2672) |
| [41] | Tät | 15 | 47 (27 - 72) | NR | 100 (15/15) | Qualitative Interview | NR |
| [42] | Tät | 123 (61/62) | 44.7 (SD 9.4) | NR | 100 (123/123) | Cost-utility analysis of a previous RCT | SUI: 100 (123/123) |
| [43] | Tät | 61 | 45 (27 - 72) | 90.2 (55/61) | 100 (61/61) | Secondary analysis of RCT treatment G | SUI: 100 (61/61) |
| [44] | Tät | 98 | 58 (37 - 77) | 86.0 (84/98) | 100 (98/98) | Secondary data analysis of RCT | MUI: 74 (72/98) and UUI: 26 (22/98) |
| [45] | Tät | 123 (61/62) | 44.7 (27-72) | 80 (98/123) | 100 (123/123) | RCT | SUI |
| [46] | MiHealth Bladder | 29 (0/29) | 54.4 (29-77) | 55 (16/29) university | 100 (29/29) | Single group pretest or posttest design | NR |



XSL•FO
RenderX

| Reference | Name of app or webpage | Sample size (control/ intervention group), % (n/N) | Mean age (range or SD) | Education level (%, university or higher) | Women (%, n/N) | Study design | IU[a] types (%, n/N) |
|---|---|---|---|---|---|---|---|
| [47] | Diário Saúde | 21 initials (9/12) | App group 47.2±10.06<br>control group 53.3±13.2 | NR | 100 (21/21) | RCT | SUI |
| [48] | Diário Saúde | 156 | 49.3±14.2 | NR | 100 (156/156) | RCT - single group | SUI |
| [49] | Yoga of Immortals YOI | 422 | (18 - 74) | NR | 94.5 (399/422) | Prospective cohort | NR |
| [50] | Web page: www.tät.nu [51] | 275 (booklet 109 and internet 166) | booklet group 59.4 years<br>Internet group 54.5 years | 59 booklet group, (64/109) and 67 internet group (111/166) | 100 (275/275) | Pragmatic prospective cohort | NR |
| [52] | iDry<br>Usage Results of a Mobile App for Managing UI[g] | 878 | 51 (SD 17.6). | NR | 25.4 (223/878) | Obser vational | NR |

[a]UI: urinary incontinence

[b]RCT: randomized controlled trial.

[c]MUI: mixed urinary incontinence.

[d]UUI: urgency urinary incontinence.

[e]NR: no results.

[f]SUI: stress urinary incontinence.

[g]UI: urinary incontinence.



XSL•FO
RenderX

**Table .** Conclusion of the papers. Incontinence episode frequency was calculated per bladder diary entries.

| Reference | ICIQ-SF | ICIQ-OB | ICIQ-LUTSqol | PGI-I | Patient satisfaction | App usage | PFMT[a] | Interview | UI[b] frequency* | Other instruments | Main results of IG[c] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| [31] | ✓ | ✓ | ✓ | ✓ | ✓ | | | | ✓ | Incontinence Catastrophizing Scale | Improvement in all parameters |
| [32] | ✓ | | ✓ | | | | | | | Costs for assessment, treatment delivery, incontinence aids, laundry, time for PFMT and BT[d], QALYs[e], and ICER[f] | Higher quality of life and annual cost (€738.42 vs €605.82). |
| [33] | ✓ | ✓ | ✓ | ✓ | ✓ | | | | | — | Reduction in the severity of UI, improvement of OB[g] symptoms, and QoL.[h] |
| [34] | ✓ | ✓ | | | | ✓ | ✓ | | | — | Long-term efficacy, achieving a significant reduction in urinary symptoms, similar in both groups. |
| [35] | ✓ | | ✓ | ✓ | | | | | ✓ | — | Long-term efficacy, achieving a significant reduction in urinary symptoms, similar in both groups. |



XSL•FO
RenderX

| Reference | ICIQ-SF | ICIQ-OB | ICIQ-LUTSqol | PGI-I | Patient satisfaction | App usage | PFMT[a] | Interview | UI[b] frequency* | Other instruments | Main results of IG[c] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| [36] | ✓ | | | | | | | | | 3IQ, eHEALS[i], TAM, and MARS[j] | The app could facilitate access to treatment, increase self-awareness, and improve adherence. |
| [37] | | | | | | | | ✓ | | — | Treatment success was influenced by factors such as adherence, personality, exercise integration, and prior experience, while the main barriers were unrealistic expectations and personal difficulties. |
| [38] | ✓ | | | | | | | ✓ | | Prognostic factors and modifiers were combined | Personal factors such as symptom severity, age, educational level, and impact on quality of life are crucial in predicting the effectiveness of eHealth treatments for UI. |



XSL•FO
RenderX

| Reference | ICIQ-SF | ICIQ-OB | ICIQ-LUTSqol | PGI-I | Patient satisfaction | App usage | PFMT[a] | Interview | UI[b] frequency* | Other instruments | Main results of IG[c] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| [39] | ✓ | | ✓ | | | | | | | EQ - 5D - 5L | Improvements in the severity of UI for at least one year, with no significant differences between groups. |
| [40] | ✓ | | | ✓ | | ✓ | | | | — | APP group: significant reduction in the severity of UI and improvement in its symptoms. |
| [41] | | | | | | ✓ | ✓ | ✓ | | — | Using the app empowered women and helped them self-manage their UI treatment by motivating them to follow a PFMT. |
| [42] | ✓ | | ✓ | | | ✓ | ✓ | | ✓ | CTA[k], training time, UI assistive devices, and laundry | The app is cost-effective for the treatment of UI. The annual cost was €547.00 app group and €482.40 control group. Greater gains in life years were observed at a similar cost. |



XSL•FO
RenderX

| Reference | ICIQ-SF | ICIQ-OB | ICIQ-LUTSqol | PGI-I | Patient satisfaction | App usage | PFMT[a] | Interview | UI[b] frequency* | Other instruments | Main results of IG[c] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| [43] | ✓ | | ✓ | ✓ | | | ✓ | | | — | Treatment expectations, weight change during treatment, and self-reported PFMS[l] muscle strength were significantly associated with successful SUI[m] treatment outcome using an app. |
| [44] | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | ✓ | — | Participants were satisfied with the treatment 15 months after having access to the app |
| [45] | ✓ | | ✓ | ✓ | | | ✓ | | | IEF | IG: improvements in symptom severity, QoL, and PGI-I scores. |
| [46] | ✓ | | | ✓ | ✓ | | ✓ | | | — | IG: It is effective in reducing UI throughout the intervention |
| [47] | ✓ | | | | | | | | | sEMG and PFM examination; ICIQ-VS; QUID. App usage. Oxford Modified Scale | Improvements in symptoms and QoL were observed, with no significant differences between groups. |



XSL•FO
RenderX

| Reference | ICIQ-SF | ICIQ-OB | ICIQ-LUTSqol | PGI-I | Patient satisfaction | App usage | PFMT[a] | Interview | UI[b] frequency* | Other instruments | Main results of IG[c] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| [48] | ✓ | ✓ | | | | | | | | QUID-Portuguese | App treatment was highly effective in improving QoL and UI |
| [49] | ✓ | | ✓ | ✓ | | | | | | — | Improvement in UI severity, QoL, and PGI-I scale at 4 and 8 weeks |
| [50] | ✓ | | | | | | | | | — | UI improvement in both groups with no significant differences |
| [52] | ✓ | ✓ | ✓ | ✓ | | | | | | — | Improvements in reducing pad use, with a 20% decrease in daily use in the short term. |

[a]PFMT: pelvic floor muscle training.
[b]UI: urinary incontinence.
[c]IG: intervention group.
[d]BT: Bladder Training
[e]QALY: quality-adjusted life-year.
[f]ICER: incremental cost-effectiveness ratio.
[g]OB: overactive bladder.
[h]QoL: quality of life.
[i]eHEALS: eHealth Literacy Scale.
[j]MARS: Mobile App Rating Scale.
[k]CTA: costs for treatment administration.
[l]PFMS: pelvic floor muscle strength.
[m]SUI: stress urinary incontinence.

The selection of outcomes in this review was guided by the primary objective of identifying mobile apps and websites that integrate BCTs and gamification elements to support bladder health self-management among adults aged 50 years and older. Accordingly, we prioritized outcomes that capture both the behavioral components of the interventions and their practical relevance for end users. For the included apps, we extracted information on cost, study setting, presence of BCTs, MARS scores [29], country of origin, and app-store availability, as these indicators reflect accessibility, design quality, and behavioral content, which are key dimensions in evaluating digital self-management tools for older adults.

No specific methods for data preparation or synthesis, such as handling of missing summary statistics or data conversions, were required.

No additional analyses were conducted to explore heterogeneity (eg, subgroup analysis or meta-regression), nor were sensitivity



XSL•FO
RenderX

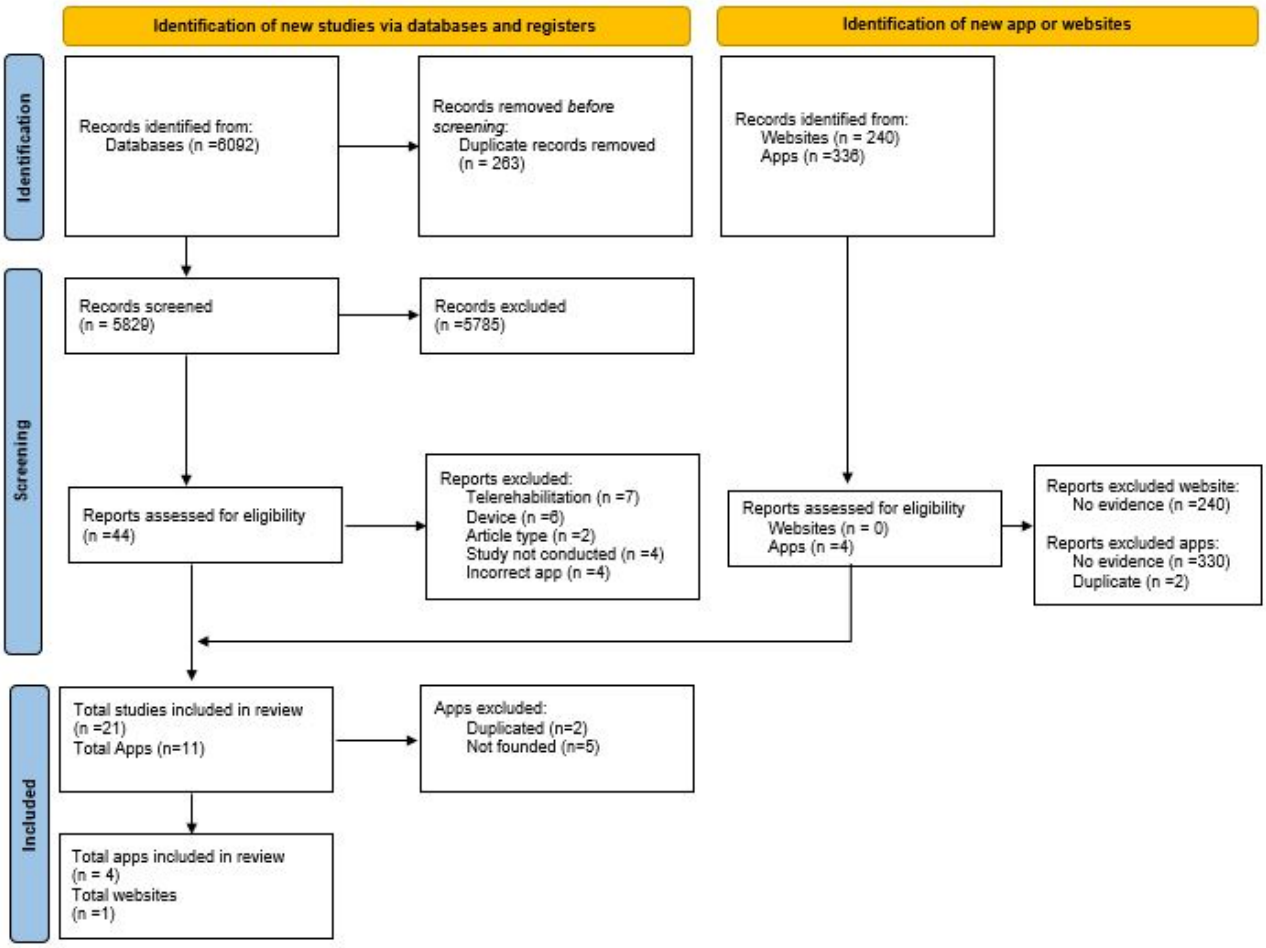
analyses performed to assess the robustness of the synthesized results. Likewise, no specific methods were applied to evaluate risk of bias due to missing results or reporting biases, and no formal assessment of certainty or confidence in the body of evidence was undertaken. These procedures were not applicable to our study.

## Results

### Overview

Following the screening process, 21 papers were deemed eligible for inclusion. These papers cover 7 distinct mobile apps and 1 website. The entire search process is illustrated in Figure 1 (flowchart).

**Figure 1.** Flowchart depicting the search and screening process.




### Database Results

From the academic database searches, we identified seven apps with scientific evidence; however, we were only able to download two of them, Tät (eContinence AB) and URinControll (Curavista). The apps that we could not download or locate were Tät II (Curavista), MyHealtheBladder (U.S. Department of Veterans Affairs), Diário Saúde (Eliza Ferrari), Yoga of Immortals–YOI (Ishan Shivanand (originating from ShivYog), and iDry (Ruslan Pavliuchenko) apps. Although these apps have scientific evidence, they were unavailable for download from Google Play or the Apple App Store.

For the app search, no apps were found in Lithuania. In Spain, we identified only 1 app (URinControl; Curavista), while in the United Kingdom, we found 3 different apps (URinControl, URinControl4all, (Curavista) and BladderBoss (Amara Therapeutics)). BladderBoss was available on both Google Play and the Apple App Store (as different versions), whereas the other apps were available in only one app store. Table 1 presents the technological and qualitative information about the apps.

The website search did not identify any platforms with scientific evidence supporting their effectiveness across the 3 countries. In total, 80 webpages were reviewed in each country, but none met the inclusion criteria. Only 1 website with scientific evidence was found through the scientific paper from the search: the Tät website (eContinence AB) [51].

Following the completion of the search process, a total of 4 mobile apps were identified as having scientific evidence and were available for download via Google Play or the Apple App Store. UrinControl4all should be considered a version of URinControl because it is based on the same scientific paper and differs only by minor variations (eg, language). We extracted data from the final papers and compiled them into Tables 2 and 3 to facilitate comparison and readability.

In the clinical questionnaire results, Tät showed the greatest reduction in UI severity (ICIQ-SF score: –3.9 points) and in



XSL•FO
RenderX

symptom-related quality of life measured using the International Consultation on Incontinence Questionnaire Lower Urinary Tract Symptoms Quality of Life. [1](ICIQ-LUTS-QoL score: –4.8 points), together with a perceived improvement of 91.7% according to the PGI-I, indicating a statistically significant improvement. URinControl or URinControl4all showed reductions of –2.16 and –4.34 points, respectively, with a perceived improvement of 65.7%, while BladderBoss assessed questionnaires specific to overactive bladder (ICIQ-OAB score: –2.13 and ICIQ-OAB-QoL score: –22.04 points) and did not report PGI-I data.

### BCT Results

In the BCT extraction, we identified that the apps had between 9 and 16 BCTs (mean 12, SD 3.27). All apps had BCTs related to goals and planning, feedback and monitoring, shaping knowledge, repetition and substitution, and comparison of outcomes. None had BCTs related to social support, regulation, antecedents, identity, scheduled consequences, or covert learning.

Regarding the distribution of BCT categories, the most frequently represented techniques were those associated with goals and planning, feedback and monitoring, shaping knowledge, repetition and substitution, and comparison of outcomes, each appearing six times, meaning that they were present in all the reviewed apps. In contrast, the techniques related to association and comparison of behavior appeared five times. Categories such as natural consequences showed a lower representation, appearing only twice, and the same frequency was observed for reward and threat, and self-belief. No BCTs were identified for social support, antecedents, regulation, identity, scheduled consequences, or covert learning. This pattern indicates that most apps emphasize self-regulatory and educational strategies rather than social or identity-based components.

### MARS Results

The MARS evaluation results were obtained through a paired review of the apps. The highest-rated app was BladderBoss (Apple App Store), which achieved a good score of 4.0 (SD 0.20). The remaining apps received scores within the acceptable range. Specifically, Tät (Spain) obtained an acceptable score of 3.39; BladderBoss on Google Play also scored 3.39 (SD 0.02); URinControl received an acceptable score of 3.33 (SD 0.07), with identical ratings in Spain and the United Kingdom; and URinControl4all achieved an acceptable score of 3.0 (SD 0.20).

This rapid review did not include detailed assessments of risk of bias for individual studies or synthesis-level evaluations of bias. Similarly, investigations of heterogeneity (such as subgroup analyses or meta-regression), sensitivity analyses to test robustness, assessments of reporting bias due to missing results, and evaluations of certainty or confidence in the body of evidence were not undertaken. These procedures were not incorporated because they fall outside the scope and objectives of a rapid review, which prioritizes timeliness and streamlined synthesis over comprehensive methodological appraisal [53].

## Discussion

### Principal Findings

A systematic search was conducted to identify mobile apps and websites related to pelvic floor training. A total of 6092 papers, 240 webpages, and 336 apps were peer reviewed, resulting in only 4 apps that were available and met our inclusion criteria. All the apps identified received an “acceptable” rating on the MARS scale [29], and one achieved a “good” rating. Regarding the BCTs, all included between 9 and 17 BCTs, and some also incorporated elements of gamification.

A total of 11 apps were identified as having scientific publications supporting their efficacy. Of these, four apps met the criteria for both scientific validation and mobile availability: URinControl, Tät, URinControl4all, and BladderBoss. All were accessible via the Google Play Store, while BladderBoss was additionally available on the Apple App Store, suggesting a dominance of android-based development in this field. The URinControl and URinControl4all apps are based on the same scientific study to validate their effectiveness. They also have similar content, differing only in language and the way users access the same program. It is important to take this into account, as URinControl4all can currently be supported by the URinControl paper [35-37]; however, if both apps are updated, it is possible that these findings may no longer be applicable.

The search was conducted in Lithuania, the United Kingdom, and Spain. Notably, no relevant apps were identified in Lithuania. In contrast, 2 apps were found in Spain and 4 in the United Kingdom. A possible explanation for this discrepancy is that all identified apps were developed in English-speaking countries, potentially eliminating the necessity for translation into other languages. Indeed, market research conducted by the Organisation for the Review of Care and Health Apps (ORCHA) has reported that of the available 311,000 health apps worldwide, 227,500 (73%) are available in the United Kingdom [54].

Regarding MARS scores [29], all assessed apps scored between 3 and 4 out of 5, indicating acceptable quality. The highest-rated app was BladderBoss (version 1.8 for iOS). The most highly rated section across all apps appears to be the “Information” section, which provides high-quality content such as text, feedback, measurements, and references. However, in some cases, the app description in the App Store and the definition of measurable objectives could be improved. However, it is important to note that app evaluation frameworks do not always reflect real-world engagement [55], with MARS scores showing to have weak correlation with user retention after downloading an app [56]. Additionally, key parameters of interest, including the implementation of gamification, and the resolution of graphics are consistently well-rated across all apps. Conversely, most apps lack sufficient visual appeal and would benefit from a more detailed description of their functionalities.

The presence of BCTs within the apps was also assessed. BladderBoss demonstrated the highest incorporation of BCTs, scoring between 16 and 17 points, depending on the evaluated version. All analyzed apps incorporated at least 7 BCTs, highlighting their potential effectiveness in facilitating



XSL•FO
RenderX

behavioral modifications. Notably, all apps included BCTs related to goals and planning, feedback and monitoring, shaping knowledge, repetition and substitution, and comparison of outcomes. These techniques are fundamental in guiding users through behavioral change processes [23]. Furthermore, association and comparison of behavior were present in 5 out of the 6 analyzed apps, indicating a widespread inclusion of these techniques.

Conversely, none of the apps integrated BCTs associated with social support, regulation, antecedents, identity, scheduled consequences, or covert learning. This absence suggests potential areas for improvement in future app development to enhance user engagement and support. A study highlights the significance of behavior change and its influence on the adoption of home-based technology by older adults [11]. Specifically, the study emphasizes how personal perceptions and behavioral factors shape technology use in this population. Given the impact of these elements on technology acceptance, future research and app development should integrate strategies that address behavior change to enhance user engagement and long-term adoption. It is also important to consider that the effectiveness of an intervention is not solely determined by the number of BCTs used, but rather by their relevance and alignment with the behavioral change needs and characteristics of the target population.

Moreover, the findings of this review suggest that the combination of BCTs and gamification elements may play a key role in supporting adherence and sustained engagement in digital interventions for UI. Core BCTs such as goal-setting, self-monitoring, feedback, and guided repetition function as psychological mechanisms that facilitate habit formation and long-term behavioral maintenance [17], while gamification elements can amplify these effects by enhancing intrinsic motivation, perceived competence, and the visibility of progress [13]. This interaction is particularly relevant for older adults, who often require additional motivational support to maintain behavioral routines over time [11]. However, the variability observed in the selection and implementation of BCTs across the included apps highlights the need for future digital interventions to adopt more explicit theoretical frameworks and a systematic selection of BCTs aligned with the behavioral needs of people with UI.

Several BCTs appear particularly relevant for UI management. Social support can reduce stigma and feelings of isolation, thereby increasing motivation and adherence to pelvic floor muscle training. Regulation strategies, such as self-monitoring of symptoms and fluid intake, enhance patients' sense of control and facilitate consistent practice. Antecedents, including the identification and modification of triggers (eg, caffeine consumption or timing of fluid intake), help minimize risk situations and improve symptom management. Strengthening identity by fostering the role of the patient as an active self-manager promotes internalization of healthy routines and long-term commitment [57]. The use of scheduled consequences, such as rewards or positive feedback for adherence to exercises, reinforces behavior through operant conditioning. Finally, covert learning techniques, including guided imagery or mental rehearsal, allow patients to visualize successful execution of exercises and coping strategies, which can increase confidence and improve real-world performance. Together, these BCTs target mechanisms of motivation, self-efficacy, and habit formation, thereby supporting adherence and sustainable behavior change in UI interventions.

The gamification elements identified in the apps were limited, although all apps include some gamification element. URinControl and Tät still follow app models with drawings instead of images, while BladderBoss used real-life pictures and videos. In all the apps, exercises are classified by levels, and the user progresses from one level to another. For example, in URinControl, users receive a medal upon completing a level.

All the apps offer the option to view the evolution or progress of the exercises and urine leakage. None incorporate a game within the app or allow users to rate the exercises, receiving a score upon completion, view user ranking, or earn rewards, among other gamification elements.

Several of the studies reviewed provided a qualitative assessment of app use through interviews. These analyses identified that adherence directly influences the outcomes of the application, acting both as a barrier (eg, difficulty in finding the time to perform the exercises) and as a facilitator. Adherence itself was shaped by personal factors, app-related characteristics, and users' awareness. A frequently reported concern was the insecurity experienced by users when performing the exercises, as many were unsure whether they were executing them correctly.

On the positive side, the novelty of technology in this type of treatment was highlighted, together with its accessibility; patients do not need to wait on long lists to receive care, and exercises are always available. The apps were also valued as practical guides that support users in performing the exercises while preserving privacy. Many participants emphasized that the use of apps increased their awareness of symptoms and improved their education about the condition.

The scientific literature consistently supports the effectiveness of mobile apps in enhancing urinary continence, improving quality of life, and increasing patients' perception of symptomatic relief. However, variability in statistical significance was observed across studies. Specifically, interventions using the Tät [32,40-42] and Tät II [31,33,34] apps demonstrated statistically significant differences between groups, whereas URinControl [35,37,39] showed improvements that were not statistically significant.

Treatment outcomes were influenced by several factors, including users' adherence, personality traits, integration of exercises into daily routines, and previous experience with pelvic floor training. Conversely, the primary barriers identified were unrealistic patient expectations and individual difficulties in maintaining adherence.

These findings highlight the necessity of optimizing digital interventions to ensure greater personalization and clinical effectiveness. Key determinants such as symptom severity, age, educational background, and the impact of UI on quality of life should be integrated into the design of eHealth interventions to enhance their predictive validity and therapeutic outcomes.



XSL•FO
RenderX

Overall, the results emphasize the importance of developing mobile apps tailored to the specific needs of the target patient population to enhance adherence, improve usability, and ultimately contribute to better symptom management and quality of life.

In summary, although few evidence-based apps or websites are available for people aged 50 and over, existing tools suggest potential benefits in reducing UI and improving quality of life. However, the mixed statistical significance across studies underscores the need for further research to confirm these effects. Moreover, this review highlights the importance of personalizing digital interventions to maximize their effectiveness and ensuring that future apps incorporate a broader range of BCTs and more robust gamification strategies. Further studies are needed to verify the effectiveness of these apps and to expand the availability of high-quality, evidence-based tools specifically designed for this population

### Limitations and Strengths

One of the main limitations of this study is that the review was conducted in only three countries, which may restrict the generalizability of our findings. Additionally, some apps were not available on the App Store, and we did not search all app markets (eg, Amazon or the Xiaomi App Store), leading to potential gaps in our evaluation. This limitation may have influenced our ability to provide a fully comprehensive analysis of all relevant apps in the field.

The study was limited by the absence of a structured instrument to objectively assess the presence of gamification in the different apps. Nevertheless, the development of such a tool holds considerable potential for future research and could contribute to more precise and standardized evaluations.

Despite these limitations, our study has several key strengths. A rigorous methodology was applied, ensuring the reliability of the findings, and the evaluation process was conducted in duplicate, involving several independent reviewers, which enhanced objectivity. Furthermore, all included apps were thoroughly assessed using the MARS [29] and BCT frameworks, providing a robust framework for evaluation. Additionally, in cases where no clear scientific evidence was available regarding an app's effectiveness, we proactively contacted the respective companies to inquire about any supporting research papers. This approach reinforced the credibility and transparency of our findings.

## Acknowledgments

AI was used only for spelling and grammatical corrections of the paper. It was not used at any stage to conduct the literature search strategy, nor for data extraction or data synthesis.

## Funding

This study is part of the KOKU Bladder project, which received funding from the Research Council of Lithuania (grant S-MIP-24-112), the Catalan Board of Physiotherapists (grant 004_2024), and the Agency for Management of University and Research Grants (AGAUR) through an Industrial PhD grant (2024 DI 00058). Javier Jerez-Roig holds a grant from the Research Council of Lithuania and the Ministry of Education, Science and Sport of the Republic of Lithuania (grant S-A-UEI-23-2).

## Data Availability

Data, analytic code, and review materials are available from the corresponding author

upon reasonable request. This includes the data collection templates, data extracted from the included studies, data used for all analyses, and other materials used in the review.

## Authors' Contributions

JJ-R and OA-G contributed equally and substantially to the conception and design of the study and to the drafting and revising of the paper. NG, MB, RS, EK, LJ, AF-V, EB, ES, and JR contributed to the conception and design of the study, reviewed the literature, websites, and apps, and contributed to the drafting and revision of the paper. All authors approved the final version of the paper for publication and agreed to be accountable for all aspects of the work, ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## Conflicts of Interest

None declared.

Checklist 1
PRISMA 2020 checklist.
[DOCX File, 51 KB - i-jmr_v15i1e76867_app1.docx ]

## References



XSL•FO
RenderX

1. Bardsley A. An overview of urinary incontinence. Br J Nurs 2016 Oct 13;25(18):S14-S21. [doi: 10.12968/bjon.2016.25.18.S14] [Medline: 27734727]
2. Milsom I, Coyne KS, Nicholson S, Kvasz M, Chen CI, Wein AJ. Global prevalence and economic burden of urgency urinary incontinence: a systematic review. Eur Urol 2014 Jan;65(1):79-95. [doi: 10.1016/j.eururo.2013.08.031] [Medline: 24007713]
3. European guidelines on urinary incontinence. World Federation of Incontinent Patients. URL: https://wfipp.org/project/digital-platform/ [accessed 2026-08-07]
4. Cornu JN, Gacci M, Hashim H, Herrmann TRW, Malde S, Netsch C, et al. EAU Guidelines on management of non-neurogenic male lower urinary tract symptoms (LUTS).: European Association of Urology; 2025. URL: https://uroweb.org/guidelines/management-of-non-neurogenic-male-luts [accessed 2026-08-10]
5. Kilpatrick KA, Paton P, Subbarayan S, et al. Non-pharmacological, non-surgical interventions for urinary incontinence in older persons: a systematic review of systematic reviews. The SENATOR project ONTOP series. Maturitas 2020 Mar;133:42-48. [doi: 10.1016/j.maturitas.2019.12.010] [Medline: 32005422]
6. Alouini S, Memic S, Couillandre A. Pelvic floor muscle training for urinary incontinence with or without biofeedback or electrostimulation in women: a systematic review. Int J Environ Res Public Health 2022 Feb 27;19(5):2789. [doi: 10.3390/ijerph19052789] [Medline: 35270480]
7. McGrother CW, Donaldson MMK, Thompson J, Wagg A, Tincello DG, Manktelow BN. Etiology of overactive bladder: a diet and lifestyle model for diabetes and obesity in older women. Neurourol Urodyn 2012 Apr;31(4):487-495. [doi: 10.1002/nau.21200] [Medline: 22374635]
8. Virtuoso JF, Mazo GZ, Menezes EC. Urinary incontinence and perineal muscle function in physically active and sedentary elderly women. Rev Bras Fisioter 2011;15(4):310-317. [doi: 10.1590/s1413-35552011005000014] [Medline: 21860992]
9. Physical activity. World Health Organization. 2024. URL: https://www.who.int/es/news-room/fact-sheets/detail/physical-activity [accessed 2026-08-07]
10. Venegas M, Carrasco B, Casas-Cordero R. Factors influencing long-term adherence to pelvic floor exercises in women with urinary incontinence. Neurourol Urodyn 2018 Mar;37(3):1120-1127. [doi: 10.1002/nau.23432] [Medline: 29095511]
11. Peek STM, Wouters EJM, van Hoof J, Luijkx KG, Boeije HR, Vrijhoef HJM. Factors influencing acceptance of technology for aging in place: a systematic review. Int J Med Inform 2014 Apr;83(4):235-248. [doi: 10.1016/j.ijmedinf.2014.01.004] [Medline: 24529817]
12. Huotari K, Hamari J. A definition for gamification: anchoring gamification in the service marketing literature. Electron Markets 2017 Feb;27(1):21-31. [doi: 10.1007/s12525-015-0212-z]
13. Ryan RM, Rigby CS, Przybylski A. The motivational pull of video games: a self-determination theory approach. Motiv Emot 2006 Dec 12;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
14. Csíkszentmihályi M. Flow: The Psychology of Optimal Experience: Harper; 1990.
15. Mazeas A, Duclos M, Pereira B, Chalabaev A. Evaluating the effectiveness of gamification on physical activity: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Jan 4;24(1):e26779. [doi: 10.2196/26779] [Medline: 34982715]
16. Selles WL, Santos EC, Romero BD, Lunardi AC. Effectiveness of gamified exercise programs on the level of physical activity in adults with chronic diseases: a systematic review. Disabil Rehabil 2024 Dec;46(26):6231-6239. [doi: 10.1080/09638288.2024.2323614] [Medline: 38444153]
17. Michie S, Richardson M, Johnston M, et al. The behaviour change technique taxonomy (v1) of 93 hierarchically clustered techniques. Ann Behav Med 2013;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
18. Chen YX, Zhou Y, Zhang XL, He WY, Ye Q, Xu M. Digital health interventions for cancer-related cognitive impairment in breast cancer patients: a scoping review. Digit Health 2025;11:20552076251358297. [doi: 10.1177/20552076251358297] [Medline: 40677519]
19. Burgio KL. Behavioral treatment options for urinary incontinence. Gastroenterology 2004 Jan;126(1 Suppl 1):S82-S89. [doi: 10.1053/j.gastro.2003.10.042] [Medline: 14978643]
20. Barnes KL, Dunivan G, Jaramillo-Huff A, Krantz T, Thompson J, Jeppson P. Evaluation of smartphone pelvic floor exercise applications using standardized scoring system. Female Pelvic Med Reconstr Surg 2019;25(4):328-335. [doi: 10.1097/SPV.0000000000000563] [Medline: 29489554]
21. Garritty C, Gartlehner G, Nussbaumer-Streit B, et al. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. J Clin Epidemiol 2021 Feb;130:13-22. [doi: 10.1016/j.jclinepi.2020.10.007] [Medline: 33068715]
22. Tricco AC, Langlois EV, Straus SE. Rapid reviews to strengthen health policy and systems: a practical guide. World Health Organization. URL: https://evidence-impact.org/storage/121/Rapid-reviews-to-strengthen-health-policy-and-systems--a-practical-guide.pdf [accessed 2026-08-07]
23. Tsertsidis A, Kolkowska E, Hedström K. Factors influencing seniors' acceptance of technology for ageing in place in the post-implementation stage: a literature review. Int J Med Inform 2019 Sep;129:324-333. [doi: 10.1016/j.ijmedinf.2019.06.027] [Medline: 31445274]




XSL•FO
RenderX

24. Gasteiger N, Dowding D, Norman G, et al. Conducting a systematic review and evaluation of commercially available mobile applications (apps) on a health-related topic: the TECH approach and a step-by-step methodological guide. BMJ Open 2023 Jun 12;13(6):e073283. [doi: 10.1136/bmjopen-2023-073283] [Medline: 37308269]
25. McGarrigle L, Boulton E, Todd C. Map the apps: a rapid review of digital approaches to support the engagement of older adults in strength and balance exercises. BMC Geriatr 2020 Nov 18;20(1):483. [doi: 10.1186/s12877-020-01880-6] [Medline: 33208117]
26. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210. [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
27. Whitehead SH, Nyman SR, Broaders F, Skelton DA, Todd CJ. The quality of English-language websites offering falls-prevention advice. Health Inform J 2012;18:50-65. [doi: 10.1177/1460458211432588]
28. Rew L, Saenz A, Walker LO. A systematic method for reviewing and analysing health information on consumer-oriented websites. J Adv Nurs 2018 May 29. [doi: 10.1111/jan.13726] [Medline: 29845648]
29. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
30. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: development and evaluation of a taxonomic method. Health Technol Assess 2015;19(99):1-188. [doi: 10.3310/hta19990]
31. Wadensten T, Nyström E, Franzén K, Lindam A, Wasteson E, Samuelsson E. A Mobile App for Self-management of Urgency and Mixed Urinary Incontinence in Women: Randomized Controlled Trial. J Med Internet Res 2021 Apr 5;23(4):e19439. [doi: 10.2196/19439] [Medline: 33818395]
32. Ekersund J, Samuelsson E, Lindholm L, Sjöström M. A mobile app for the treatment of female mixed and urgency incontinence: a cost-effectiveness analysis in Sweden. Int Urogynecol J 2022 May;33(5):1273-1282. [doi: 10.1007/s00192-022-05137-1] [Medline: 35278093]
33. Wadensten T, Nyström E, Nord A, Lindam A, Sjöström M, Samuelsson E. App-based self-management of urgency and mixed urinary incontinence in women: One-year follow-up. Neurourol Urodyn 2022 Apr;41(4):945-954. [doi: 10.1002/nau.24898] [Medline: 35266189]
34. Hoffman V, Söderström L, Samuelsson E. Self-management of stress urinary incontinence via a mobile app: two-year follow-up of a randomized controlled trial. Acta Obstet Gynecol Scand 2017 Oct;96(10):1180-1187. [doi: 10.1111/aogs.13192] [Medline: 28718223]
35. Loohuis AMM, Wessels NJ, Dekker JH, et al. App-based treatment in primary care for urinary incontinence: a pragmatic, randomized controlled trial. Ann Fam Med 2021;19(2):102-109. [doi: 10.1370/afm.2585] [Medline: 33685871]
36. Wessels NJ, Hulshof L, Loohuis AMM, et al. User experiences and preferences regarding an app for the treatment of urinary incontinence in adult women: qualitative study. JMIR Mhealth Uhealth 2020 Jun 12;8(6):e17114. [doi: 10.2196/17114] [Medline: 32530431]
37. Wessels NJ, Loohuis AMM, van der Worp H, et al. Barriers and facilitators associated with app-based treatment for female urinary incontinence: mixed methods evaluation. JMIR Mhealth Uhealth 2021 Sep 17;9(9):e25878. [doi: 10.2196/25878] [Medline: 34533466]
38. Loohuis AMM, Burger H, Wessels N, et al. Prediction model study focusing on eHealth in the management of urinary incontinence: the Personalised Advantage Index as a decision-making aid. BMJ Open 2022 Jul 25;12(7):e051827. [doi: 10.1136/bmjopen-2021-051827] [Medline: 35879013]
39. Loohuis AMM, Van Der Worp H, Wessels NJ, et al. Cost-effectiveness of an app-based treatment for urinary incontinence in comparison with care-as-usual in Dutch general practice: a pragmatic randomised controlled trial over 12 months. BJOG 2022 Aug;129(9):1538-1545. [doi: 10.1111/1471-0528.17191] [Medline: 35460163]
40. Rygh P, Asklund I, Samuelsson E. Real-world effectiveness of app-based treatment for urinary incontinence: a cohort study. BMJ Open 2021 Jan 4;11(1):e040819. [doi: 10.1136/bmjopen-2020-040819] [Medline: 33397664]
41. Asklund I, Samuelsson E, Hamberg K, Umefjord G, Sjöström M. User experience of an app-based treatment for stress urinary incontinence: qualitative interview study. J Med Internet Res 2019 Mar 14;21(3):e11296. [doi: 10.2196/11296] [Medline: 30869644]
42. Sjöström M, Lindholm L, Samuelsson E. Mobile app for treatment of stress urinary incontinence: a cost-effectiveness analysis. J Med Internet Res 2017 May 8;19(5):e154. [doi: 10.2196/jmir.7383] [Medline: 28483745]
43. Nyström E, Asklund I, Sjöström M, Stenlund H, Samuelsson E. Treatment of stress urinary incontinence with a mobile app: factors associated with success. Int Urogynecol J 2018 Sep;29(9):1325-1333. [doi: 10.1007/s00192-017-3514-1] [Medline: 29222718]
44. Wadensten T, Nyström E, Sjöström M, Lindam A, Samuelsson E. APP-based treatment of urgency and mixed urinary incontinence in women: factors associated with long-term satisfaction. Arch Gynecol Obstet 2024 May;309(5):2193-2202. [doi: 10.1007/s00404-023-07303-2] [Medline: 38141064]
45. Asklund I, Nyström E, Sjöström M, Umefjord G, Stenlund H, Samuelsson E. Mobile app for treatment of stress urinary incontinence: a randomized controlled trial. Neurourol Urodyn 2017 Jun;36(5):1369-1376. [doi: 10.1002/nau.23116] [Medline: 27611958]




XSL•FO
RenderX

46. Goode PS, Markland AD, Echt KV, et al. A mobile telehealth program for behavioral treatment of urinary incontinence in women veterans: development and pilot evaluation of MyHealtheBladder. Neurourol Urodyn 2020 Jan;39(1):432-439. [doi: 10.1002/nau.24226] [Medline: 31774200]
47. Araujo CC, Marques ADA, Juliato CRT. The adherence of home pelvic floor muscles training using a mobile device application for women with urinary incontinence: a randomized controlled trial. Female Pelvic Med Reconstr Surg 2020 Nov;26(11):697-703. [doi: 10.1097/SPV.0000000000000670] [Medline: 30624250]
48. Araujo CCD, Brito LGO, Marques A, Bardin M, Juliato CRT. Mobile app for pelvic floor muscle training for urinary incontinence during the coronavirus disease 2019 pandemic: clinical trial. Rev Assoc Med Bras (1992) 2024;70(3):e20231073. [doi: 10.1590/1806-9282.20231073] [Medline: 38656000]
49. Tunuguntla H, Tunuguntla R, Kathuria H, et al. App-based yoga of immortals: a novel, easy-to-use intervention in the management of urinary incontinence. Urology 2022 Sep;167:73-81. [doi: 10.1016/j.urology.2022.06.011] [Medline: 35788018]
50. Bokne K, Sjöström M, Samuelsson E. Self-management of stress urinary incontinence: effectiveness of two treatment programmes focused on pelvic floor muscle training, one booklet and one Internet-based. Scand J Prim Health Care 2019 Sep;37(3):380-387. [doi: 10.1080/02813432.2019.1640921] [Medline: 31317808]
51. Tät.nu – research into the treatment of urinary incontinence. TäT.nu. URL: https://tät.nu/language/en/english-start/ [accessed 2026-08-07]
52. Pepper J, Zhang A, Li R, Wang XH. Usage results of a mobile app for managing urinary incontinence. J Urol 2015 Apr;193(4):1292-1297. [doi: 10.1016/j.juro.2014.10.009] [Medline: 25308623]
53. Tricco AC, Antony J, Zarin W, et al. A scoping review of rapid review methods. BMC Med 2015 Sep 16;13(224):224. [doi: 10.1186/s12916-015-0465-6] [Medline: 26377409]
54. Market facts and figures: the digital health market. 2022. URL: https://orchahealth.com/market-facts-and-figures [accessed 2026-08-07]
55. Alon N, Torous J. Current challenges for evaluating mobile health applications. J Am Med Inform Assoc 2023 Feb 16;30(3):617-624. [doi: 10.1093/jamia/ocac244] [Medline: 36484621]
56. Kaveladze BT, Wasil AR, Bunyi JB, Ramirez V, Schueller SM. User experience, engagement, and popularity in mental health apps: secondary analysis of app analytics and expert app reviews. JMIR Hum Factors 2022 Jan 31;9(1):e30766. [doi: 10.2196/30766] [Medline: 35099398]
57. Kenne K, Bradley CS. Behavioral therapy and lifestyle modifications for the management of urinary incontinence in women. In: Cameron AP, editor. Female Urinary Incontinence: Springer; 2022:1-20. [doi: 10.1007/978-3-030-84352-6_6]


## Abbreviations

**BCT:** behavior change technique
**ICIQ-OAB:** International Consultation on Incontinence Questionnaire - Overactive Bladder
**ICIQ-SF:** International Consultation on Incontinence Questionnaire - Short Form
**MARS:** Mobile Application Rating Scale
**ORCHA:** Organisation for the Review of Care and Health Apps
**PA:** physical activity
**PGI-I:** Patient Global Impression of Improvement
**PICO:** Population, Intervention, Comparison, and Outcome
**PMPT:** Pelvic Muscle Physical Therapy
**UI:** urinary incontinence


*Edited by M Balcarras; submitted 02.May.2025; peer-reviewed by A Mazeas, B Sangelaji; revised version received 02.Jun.2026; accepted 25.Jun.2026; published 26.Aug.2026.*

*Please cite as:*
*Aguila-Gimeno O, Gasteiger N, Stanmore E, Brazaitis M, Solianik R, Karkauskiene E, Jarutienė L, Romeu-Busquets J, Jerez-Roig J Digital Health Interventions Incorporating Behavior Change Techniques and Gamification for Bladder Health in Adults Aged 50 Years and Older: Rapid Review*
*Interact J Med Res 2026;15:e76867*
*URL: https://www.i-jmr.org/2026/1/e76867*
*doi:10.2196/76867*


© Oscar Aguila-Gimeno, Norina Gasteiger, Emma Stanmore, Marius Brazaitis, Rima Solianik, Erika Karkauskiene, Laura Jarutienė, Júlia Romeu-Busquets, Javier Jerez-Roig. Originally published in the Interactive Journal of Medical Research



XSL•FO
RenderX

(https://www.i-jmr.org/), 26.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Health Information Technologies in the Support System of New Fathers' Postnatal Depression: Scoping Review

Weijie Wang[1], MDes; Erika Penney[2], DClinPsy; Jaime Garcia[1], PhD

[1]Faculty of Engineering and Information Technology, University of Technology Sydney, 15 Broadway, Ultimo, New South Wales, Australia
[2]Faculty of Health, University of Technology Sydney, Ultimo, New South Wales, Australia

**Corresponding Author:**
Jaime Garcia, PhD
Faculty of Engineering and Information Technology, University of Technology Sydney, 15 Broadway, Ultimo, New South Wales, Australia

## Abstract

**Background:** Perinatal mental health issues are widely recognized as a significant public health concern; however, fathers remain underrecognized and underserved within perinatal care systems. Existing reviews have largely focused on mothers, infants, couples, or general paternal mental health interventions, leaving limited synthesis of technology-based and digitally delivered support relevant to paternal perinatal depression. Digital interventions, including SMS programs, online groups, web-based platforms, mobile health apps, and hybrid digital follow-up, may offer flexible ways of engaging fathers; however, their evidence base remains unclear.

**Objective:** This scoping review aimed to map the available evidence on technology-based, digitally delivered, and hybrid interventions relevant to paternal postnatal depression and paternal perinatal mental health. Specifically, it examined the types of evidence available; how interventions varied by delivery mode, methodology, and population; how interventions were characterized in terms of format and duration; and what outcomes had been measured across studies.

**Methods:** This review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) framework. A comprehensive search was conducted between February and April 2025 and updated in June 2026 across multidisciplinary electronic databases. Search terms covered 4 concepts: fathers or nonbirthing parents, perinatal mental health, digital or technology-based modalities, and intervention or evaluation purpose. Eligible studies were peer-reviewed English-language articles in which fathers, expectant fathers, new fathers, male partners, or nonbirthing parents were active participants, target users, or reported subgroups, and in which a digital, technology-based, or hybrid support strategy was described or evaluated. Data were charted and synthesized descriptively and comparatively.

**Results:** Six studies met the inclusion criteria. The evidence included a feasibility study of an SMS-based, father-specific program, randomized controlled trials of hybrid psychoeducational or paternal support interventions, a pilot randomized feasibility trial of an online cognitive behavioral therapy–based intervention for pregnant people and partners, a mixed methods online group intervention for new fathers, and a feasibility, acceptability, and usability study of a father-specific mobile health app. Delivery modes included SMS messaging, WhatsApp follow-up, online video, telephone follow-up, online groups, web-based intervention content, and mobile app features. Outcomes included depression, anxiety, stress, parenting stress, dyadic adjustment, marital quality, resilience, usability, acceptability, engagement, adherence, and qualitative user experience. Findings were most consistent for feasibility, acceptability, and engagement, while evidence for effectiveness in reducing paternal postnatal depression remained limited and mixed.

**Conclusions:** The current evidence base is small, heterogeneous, and preliminary. Digital interventions may offer acceptable ways to engage some fathers and partners during the perinatal period. However, there is insufficient evidence to draw firm conclusions about effectiveness, scalability, or suitability for fathers experiencing perinatal depression. Future research should prioritize father-specific, theory-informed, co-designed digital interventions using validated paternal outcomes, longer follow-up, transparent engagement reporting, and more diverse samples.

***(Interact J Med Res 2026;15:e95238)*** doi:10.2196/95238

**KEYWORDS**

paternal postnatal depression; paternal mental health; digital intervention; digital health; perinatal education



XSL•FO
RenderX

## Introduction

The transition to fatherhood can be stressful and overwhelming. Along with the usual difficulties of new caregiver responsibilities, new fathers may face changes to their self-identities, their connection to their partners, and financial commitments [1]. Approximately 10% of new fathers experience depression or anxiety in the perinatal period, and paternal depression has been associated with reduced father involvement and poorer child development outcomes. Decreased paternal engagement in child-rearing can also lead to an increase in children with psychiatric disorders, behavioral issues, and lower academic achievement [2]. Despite this, most perinatal support is received by mothers and infants. Father-specific support is limited, and fathers are often treated as visitors before, during, and after the birth of their baby [3]. Due to stigma and societal perspectives on masculinity, fathers can tend to see themselves in a traditional provider, supporter, and caregiver role, which may create an invisible barrier to seeking emotional support from others [4].

Importantly, unlike the typical signs of mothers' depression, such as sadness, tearfulness, and withdrawal, new fathers are more likely to present with exhaustion, difficulty focusing, and irritability [5]. These gender differences may contribute to systematic underrecognition of paternal mental health problems. The lack of support can escalate challenges into more acute mental health issues, and fathers' mental well-being can impact the well-being of partners and children [6].

Existing reviews demonstrate the broader problem but do not yet provide a clear map of appropriate support for fathers' perinatal depression. Reviews of partner-inclusive or paternal perinatal interventions have found promising effects, but studies often involve mothers, infants, couples, or general parenting support rather than interventions designed primarily for the unique needs of fathers [7-10]. Reviews of paternal perinatal depression and anxiety consistently identify limited father-specific evidence, inconsistent outcome measures, and methodological weaknesses, but have stopped short of proposing a research agenda to overcome them [1,10]. While fathers have been included in some technology-based parenting programs, relatively few have focused on fathers as participants or as primary users of perinatal mental health support. This is significant because technology-based interventions may be well-suited to address the access barriers of fathers.

Digital and technology-based interventions may offer a relevant pathway for engaging fathers during the perinatal period; however, their efficacy and acceptability for this population remain understudied. Digital or technology-based interventions refer to interventions, support strategies, or service pathways in which information or communication technologies are used to deliver, reinforce, monitor, or facilitate mental health support. These may include messaging platforms, mobile health (mHealth) apps, web-based platforms, online groups, online videos, telephone or app-based follow-up, and gamified or game-informed tools.

These modalities may be particularly relevant, given that research has identified barriers to paternal help-seeking, including stigma, masculine norms, limited father-specific services, work and caregiving commitments, and reluctance to engage with conventional face-to-face support [4], highlighting the need for novel interventions to reach this population. Digital technologies may offer this novel, flexible, and private access to support [11], such as father-specific information aligned with perinatal milestones, self-care needs, supporting a partner, and caring for an infant. While digital parenting and online psychological interventions have been increasingly explored in perinatal contexts, fathers are often included as secondary participants rather than primary users [12,13]. Father-specific digital interventions may therefore provide flexible, private, timely, or context-specific support.

Existing perinatal reviews have focused almost exclusively on maternal and infant mental health, systematically excluding fathers as a population of concern. This leaves 2 critical gaps in existing literature. First, fathers are routinely excluded from perinatal interventions, which are primarily designed around mothers' experiences and needs, resulting in a scarcity of interventions specifically targeting paternal perinatal mental health. Second, the potential of digital delivery modalities to reach this underserved population remains unexplored and unmapped. In this review, the perinatal period is understood as the period from pregnancy through to the first 12 months after birth, which is commonly used in perinatal mental health research and reporting [14], and reflects the psychological continuity between pregnancy, birth, and the postpartum year.

This review aims to map and chart the extent of available evidence for technology-based and digitally delivered interventions for paternal perinatal mental health. The review addresses 3 research questions (RQs). First, what types of evidence exist and how do studies vary by delivery mode, methodology, and population? Second, how are the interventions characterized in terms of delivery mode, format, durations, and intervention components? Third, what outcomes have been measured across studies and what gaps remain in the evidence base?

## Methods

### Overview

This scoping review was conducted in accordance with the latest PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines (Checklist 1) [15]. The PRISMA-ScR checklist includes 20 essential reporting items and 2 optional items designed to ensure transparency and rigor in scoping review methodology. A comprehensive literature search was conducted between February and April 2025, and updated in June 2026, across 9 electronic databases. The screening process was conducted in 3 stages: title screening, abstract screening, and full-text screening. If eligibility could not be confidently determined at the title or abstract stage, the paper was advanced to the next stage for further review. Papers were assessed to ensure that they met all 5 inclusion criteria. Discrepancies or uncertainties during screening were resolved in discussion with a third reviewer, and consensus was reached in all cases after this discussion.



XSL•FO
RenderX

## Protocol and Registration

This scoping review was conducted in accordance with the PRISMA-ScR framework. A protocol for this review was not prospectively registered; however, the review questions, eligibility criteria, search concepts, and data-charting categories were developed before data extraction to minimize bias.

## Search Strategy and Study Selection

A scoping review format was selected in response to the limited research literature on digital interventions for paternal perinatal mental health. This approach enabled the timely integration of current evidence, highlighted gaps in the literature, and identified key areas for future research.

The search was conducted between February and April 2025 across 9 electronic databases, including technical databases (eg, ACM and IEEE Xplore), Scopus databases (eg, MEDLINE and PubMed), ScienceDirect, SpringerLink, and Wiley. An updated search was conducted in June 2026 to capture recent literature, including an additional search of the PsycINFO database. The search strategy was organized around 4 core concepts: population, mental health or well-being, technology modality, and intervention or evaluation purpose. Search terms included father-related terms, such as father*, paternal, new dad, expecting father, father-to-be, and non-birthing parent*; mental health and well-being terms, such as postnatal depression (PND), postpartum depression, perinatal mental health, anxiety, stress, psychological well-being, adjustment, help-seeking, and psychosocial support; technology terms, such as digital, online, web-based, internet, eHealth, mHealth, mobile app*, smartphone, telehealth, gamif*, game*; and intervention or evaluation terms, such as intervention, program, therap*, treatment, support, education, psychoeducation, feasibility, acceptability, usability, implementation, and evaluation. The query terms used in the search strategy are presented in Textbox 1.

An example search string was (father* OR paternal OR “new dad*” OR “expecting father*” OR “father-to-be” OR “male partner*” OR “non-birthing parent*”) AND (“postnatal depression” OR “postpartum depression” OR “perinatal mental health” OR anxiety OR stress OR wellbeing OR “parenting stress” OR adjustment OR “help-seeking”) AND (digital OR online OR web-based OR internet OR eHealth OR mHealth OR “mobile app*” OR smartphone OR SMS OR “text message*” OR WhatsApp OR telehealth OR gamif* OR game* OR “serious game*” OR “game-based”) AND (intervention* OR program* OR therap* OR treatment OR support OR education OR psychoeducation OR feasibility OR acceptability OR usability OR implementation OR engagement OR evaluation).

These terms were searched across titles, abstracts, keywords, and full texts. The search string was adapted to the syntax and functionality of each database. To enhance the search strategy and refine the RQs, faculty librarians with expertise in systematic and scoping review methodology were consulted.



XSL•FO
RenderX

**Textbox 1.** Summary and search terms.

Population mental health and well-being

- Father*
- Paternal
- New dad
- Expecting father
- Father to be
- Nonbirthing parents
- Postnatal depression
- Postpartum depression
- Prenatal mental health
- Anxiety
- Stress
- Adjustment
- Psychological well-being
- Help-seeking
- Psychological support

Technology terms

- Digital
- Online
- Web-based
- Website
- Internet
- eHealth
- mHealth
- Mobile app*
- Smartphone
- Telehealth
- Gamif*
- Game*

Intervention or evaluation

- Intervention
- Program
- Therap*
- Treatment
- Support
- Education
- Psychoeducation
- Feasibility
- Acceptability
- Usability
- Implementation
- Evaluation



XSL•FO
RenderX

### Eligibility Criteria and Screening Process

This review focused on technology-based interventions targeting paternal perinatal mental health, and only papers that met the predefined inclusion and exclusion criteria were retained for analysis, as detailed in Textbox 2.

**Textbox 2.** Inclusion and exclusion criteria.

Inclusion criteria

- Full-text, peer-reviewed studies published in English.
- Fathers, expectant fathers, new fathers, male partners, or nonbirthing parents were active participants, target users, or explicitly reported as a relevant subgroup, including parent or couple interventions, where paternal or partner outcomes were reported.
- Studies addressed paternal PND, perinatal mental health, psychological well-being, help-seeking, parenting stress, adjustment to fatherhood, partner support, or closely related outcomes during the perinatal period.
- Studies described a digital, technology-based service pathway or support strategy.
- Studies reported intervention design, protocol details, feasibility, usability, acceptability, implementation, engagement, screening, referral, or relevant outcome data.

Exclusion criteria

- Fathers or nonbirthing parents were not active participants, target users, or a reported subgroup.
- Paternal mental health or well-being was not targeted or meaningfully addressed.
- Studies were not peer-reviewed or not available in English.
- The intervention was primarily focused on clinical simulation or professional training rather than support for fathers in community or home contexts.

A total of 3223 records were obtained from the search strategy listed in Textbox 1. All records were initially screened for duplicates by author 1, resulting in 82.2% (2649/3223) unique papers. Author 1 then screened all titles using the inclusion criteria, identifying 1.2% (32/2649) potentially relevant papers for the review. Because only 32 papers remained after title screening, the review team proceeded directly to full-text screening rather than conducting a separate abstract screening stage, allowing for a more thorough assessment of eligibility. Full texts were independently assessed by 2 reviewers against the predefined inclusion criteria, with author 1 identifying 11 papers for potential inclusion and author 2 identifying 6 (reviewer agreement 84%, Cohen [16] $\kappa$=0.61). Discrepancies were resolved through discussion between the 2 reviewers and arbitration by author 3, who has extensive experience in digital technologies in health application research.

Although initial agreement was moderate, both reviewers agreed on all 6 studies included. The discrepancies arose because author 1, newer to review methodology, identified 5 additional studies that did not meet full criteria. Following discussion with author 2 and author 3, who have more experience in digital health research, consensus was reached that these additional studies did not meet eligibility criteria, resulting in full agreement (100%) on the final 6 included studies.

### Data Extraction and Analysis

The following outlines the data extraction and analysis procedures for each RQ, corresponding to the structure of the *Results* section. Data were charted using a structured extraction form developed for this review. Extracted fields included author and year, country, study design, evidence type, population, sample size, intervention description, delivery mode, delivery format, delivery duration, digital component, theoretical or design basis, outcome measures, data collection method, key feasibility, acceptability, usability, engagement, or outcome findings. Data extraction was conducted by the first author, with uncertainties and interpretive decisions discussed collaboratively with the research team.

Findings were synthesized using a descriptive and comparative approach that focused on mapping the type and maturity of evidence, comparing delivery modes and formats, identifying how fathers were positioned within each study, summarizing measured outcomes, and identifying gaps in the evidence base. Consistent with the exploratory aims of a scoping review, and given the small number of studies and substantial heterogeneity across interventions, populations, and outcomes, a systematic review or meta-analysis was not considered appropriate for this evidence base.

## Results

Six studies met the eligibility criteria for this scoping review. They varied in design and methodological purposes, reflecting a heterogeneous evidence base characteristic of an early-stage, emerging field of research. Studies were published between 2017 and 2025 and were conducted across Australia, the United States, the Netherlands, Turkey, and Israel. Table 1 presents the characteristics of each study, and Figure 1 details the specific numbers from each database.



XSL•FO
RenderX

**Table .** Overview of the included studies.

| Author and year | Fletcher et al [17] (2017) | Missler et al [18] (2020) | Canfield et al [19] (2023) | Karakoç and Kalkan [20] (2024) | Kaner et al [21] (2024) | Teague et al [11] (2025) |
|---|---|---|---|---|---|---|
| Country | Australia | Netherlands | United States | Turkey | Israel | Australia |
| Delivery mode | Smartphone-based SMS/text-message intervention | Hybrid psychoeducational intervention with online and digital components | Online web-based intervention | Hybrid postpartum education+digital follow-up | Online group intervention | Mobile health app intervention |
| Delivery format | Short SMS messages with embedded links and mood-tracking questions | Booklet, online video, prenatal home visit, and postnatal telephone call | Self-administered online CBT[a]-based Mothers and Babies Online Course, delivered through Qualtrics | Paternal support training delivered before postpartum discharge, with WhatsApp follow-up | Father-only online discussion groups led by a male facilitator | Prototype app, Rover, including chatbot, mood tracking, goal tracking, CBT/mindfulness exercises, resources, notifications, and gamification |
| Delivery duration | From 12 weeks of gestation to 24 weeks after birth | Delivered during late pregnancy and early postpartum, intervention included prenatal components and a postnatal phone call | 8-week course; one lesson per week recommended | Initial training within the first 48 hours postpartum; follow-up to postpartum day 10 | Six weekly sessions, each approximately 75 minutes | Fathers used the app for 4 weeks; clinicians evaluated it for 1 week |
| Methodology | Feasibility and acceptability study | Randomized controlled trial | Mixed methods pilot randomized controlled feasibility trial | Randomized controlled trial | Mixed methods randomized controlled trial | Feasibility, acceptability, and usability study using a design science approach |
| Population | Soon-to-be and new fathers | Pregnant women and partners/expecting parents | Pregnant people with elevated depression or anxiety symptoms and their cohabiting partners | Married postpartum couples/fathers and mothers | New partnered fathers with an infant under 1 year | Fathers with pregnant partners or infants under 12 months; mental health clinicians |
| Sample size | 520 enrolled fathers | 138 pregnant women and 96 partners | 30 dads; 15 intervention and 15 controls | 80 fathers randomized; data collected from 160 couples | 122 fathers; 62 intervention and 60 comparison | 43 fathers and 10 mental health clinicians |
| Intervention focus | Supporting fathers' transition to fatherhood through father-infant connection, partner support, and fathers' self-care | Preventing postpartum parenting stress, depression, and anxiety, improving parental well-being and caregiving quality | Reducing perinatal mood and anxiety symptoms through online CBT-based psychoeducation and partner-inclusive support | Increasing paternal support and improving dyadic adjustment, happiness, and psychological resilience | Supporting fatherhood transition, normalizing experiences, reducing loneliness, increasing awareness of fatherhood role, and supporting couple relationship | Father-specific perinatal mental health support through mood tracking, mindfulness, goal setting, CBT-informed exercises, chatbot support, and gamified app design |



XSL•FO
RenderX

| Author and year | Fletcher et al [17] (2017) | Missler et al [18] (2020) | Canfield et al [19] (2023) | Karakoç and Kalkan [20] (2024) | Kaner et al [21] (2024) | Teague et al [11] (2025) |
|---|---|---|---|---|---|---|
| Outcome measures used | K6[b,c], completion/retention, link-click engagement, mood-tracker responses, and post-study feedback survey | Parenting Stress Index items[b], EPDS[b,d], HADS[e] anxiety[b], parental well-being, parenting self-efficacy, satisfaction with parenting role, sleep, parent-infant bonding, breastfeeding duration, room-sharing, infant crying/feeding/sleeping problems, and intervention usefulness | PHQ-9[b,f] for eligibility/screening, GAD-7[b,g], EPDS[b], retention, adherence analytics, satisfaction, and acceptability interviews | Revised Dyadic Adjustment Scale[b], Short Form Oxford Happiness Questionnaire[b], and Brief Resilience Scale[b] | EPDS[b], Israeli Marital Quality Scale[b], attendance rate, session transcripts, and 2-year follow-up feedback questionnaire | EPDS[b], DASS-21[h] anxiety and stress subscales[b], System Usability Scale[b]; Mobile App Rating Scale[b], Mobile Agnew Relationship Measure[b], app usage statistics, and qualitative feedback |
| How outcomes were collected | Automated program data, including completion, link clicks, and mood-tracker responses; K6 screening; post-study survey | Online questionnaires sent to mothers and fathers, intervention uptake/usefulness ratings, and postpartum follow-up assessments | Online surveys at baseline, 4 weeks, and 8 weeks; eMB[i] platform analytics; postintervention interviews via Zoom (Zoom Communications, Inc) | Self-report questionnaires collected from mothers and fathers at pretest within 48 hours after birth and posttest on postpartum day 10 | Quantitative self-report questionnaires, attendance records, transcripts of closing group sessions, and email feedback questionnaire 2 years after the intervention | Baseline and follow-up surveys, app usage analytics, validated usability/app-quality/therapeutic-alliance scales, and open-ended qualitative feedback |

[a]CBT: cognitive behavioral therapy.

[b]Validated measure or validated scale/subscale.

[c]K6: Kessler Psychological Distress Scale-6.

[d]EPDS: Edinburgh Postnatal Depression Scale.

[e]HADS: Hospital Anxiety and Depression Scale.

[f]PHQ-9: Patient Health Questionnaire-9.

[g]GAD-7: Generalized Anxiety Disorder-7.

[h]DASS-21: Depression Anxiety Stress Scales-21.

[i]eMB: electronic mothers and babies course.

The first RQ examined the types of evidence available and how studies varied by delivery mode, methodology, and population. Delivery modes varied substantially, including SMS messaging delivered to fathers' smartphones [17], WhatsApp-supported follow-up [20], online video within a multicomponent intervention [18], an asynchronous online cognitive behavioral therapy (CBT)–based course [19], synchronous online father-only groups [21], and a prototype mHealth app, called the Rover app [11]. Methodologically, the studies varied in the extent to which digital delivery was central to the intervention. In the SMS4dads feasibility study [17], the digital component was the main intervention mechanism, delivered through text messages to fathers' smartphones. Similarly, the Rover app study [11] placed digital design at the center of the intervention by evaluating a father-specific mHealth app. The online group intervention [21] and online CBT-based dyadic intervention [19] also used digital delivery as the primary access pathway. In contrast, the psychoeducational intervention [18] and paternal support training [20] used digital elements as supplementary components, such as online video, telephone follow-up, or WhatsApp follow-up. The target populations also varied. Several studies were explicitly father-focused, including SMS4dads, the online father group intervention, and the Rover mHealth app [11,17,21], while other studies included fathers or partners within couple, dyadic, or expectant-parent interventions [18-20].

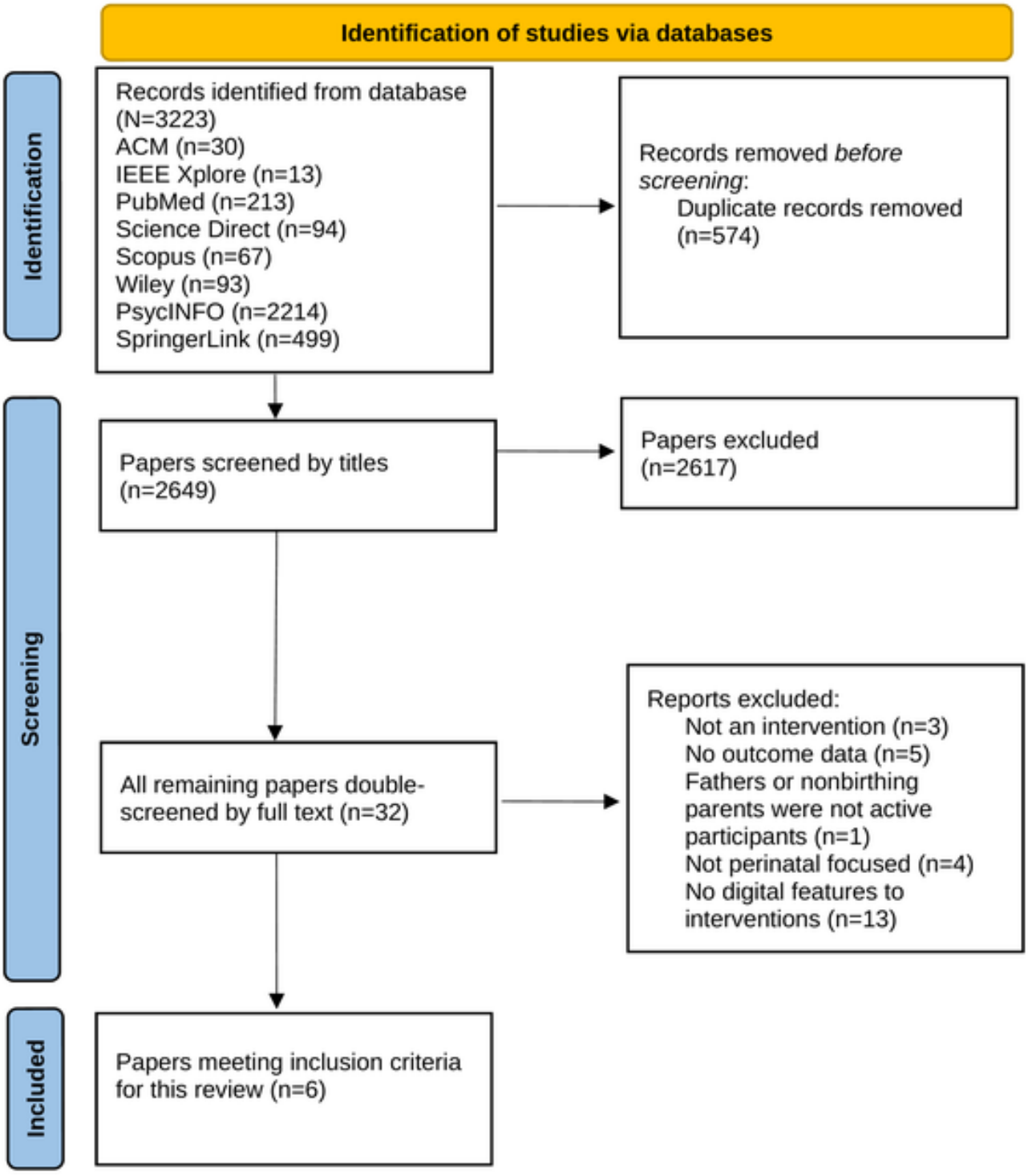
The second RQ examined how the interventions were characterized in terms of delivery mode, format, duration, and intervention components. The interventions can be grouped into 3 delivery models: fully digital, online group-based, and hybrid interventions with digital reinforcement. The fully digital interventions, SMS4dads [17] and the Rover mHealth app [11], were accessed independently of face-to-face health services. SMS4dads delivered stage-based text messages and embedded links across pregnancy and early fatherhood, while Rover offered app-based features such as mood tracking, goal tracking, mindfulness exercises, chatbot interaction, resources, and gamified elements.

The online group [21] and online CBT-based [19] intervention used digital delivery to create structured support spaces. The former comprised 6 weekly father-only group sessions led by a male facilitator, focused on peer normalization and fatherhood identity; the latter was an 8-week structured course for pregnant people and partners targeting perinatal mood and anxiety symptoms.



XSL•FO
RenderX

**Figure 1.** Screening and inclusion process.




The hybrid interventions [18,20] used digital components to reinforce nondigital support. The psychoeducational intervention [18] combined a booklet, online video, prenatal home visit, and postnatal telephone call, while the paternal support training [20] included WhatsApp follow-up after an initial in-person session.

Duration also varied: SMS4dads [17] spanned from pregnancy to postpartum; Rover [11] was evaluated over 4 weeks, the online group intervention [21] ran 6 weeks, the CBT course [19] ran 8 weeks, and the hybrid interventions [18,20] involved shorter, more targeted contact around late pregnancy and early postpartum.

The third RQ examined what outcomes have been measured across studies and what gaps remain in the evidence base. All 6 studies reported completed intervention, feasibility, acceptability, usability, or outcome evaluations [11,17-21]. Across the included studies, outcome measurement clustered around 4 broad domains: mental health symptoms, relational or dyadic functioning, feasibility and engagement, and acceptability or usability. Mental health symptoms were the most commonly measured domain. Five studies included measures of psychological distress, depression, anxiety, stress, or parenting stress [11,17-19,21]. The Edinburgh Postnatal



XSL•FO
RenderX

Depression Scale (EPDS) was the most frequently used depression-related measure, appearing across several studies. Other symptom measures included the Kessler Psychological Distress Scale-6, Patient Health Questionnaire-9, Generalized Anxiety Disorder-7, Hospital Anxiety and Depression Scale anxiety subscale, and Depression Anxiety Stress Scales-21 anxiety and stress subscales.

Relational, dyadic, and family-functioning outcomes were also represented, but they were measured inconsistently. Some studies assessed dyadic adjustment, marital quality, partner relationship quality, happiness, psychological resilience, parenting self-efficacy, parent-infant bonding, or caregiving quality [18,20,21].

Feasibility and engagement were measured in several studies [11,17,19,21], particularly those evaluating digital or app-based interventions. SMS4dads reported completion, retention, link-click engagement, and responses to mood-tracking questions. The online CBT-based intervention collected adherence data through platform analytics, while the Rover mHealth app collected app usage statistics. The online father group intervention reported attendance rates, and other studies collected data on intervention usefulness or feedback.

Acceptability, usability, and user experience were also commonly examined, but mainly in studies where the digital component was central to the intervention. The Rover app study used validated usability and app-quality measures, including the System Usability Scale, Mobile App Rating Scale, and Mobile Agnew Relationship Measure, alongside qualitative feedback [11]. SMS4dads [17] and the online CBT-based intervention [19] also collected postintervention feedback or acceptability data. Other hybrid studies assessed perceived usefulness or qualitative experiences [18,20,21].

Several gaps were evident in outcome measurement across the included studies. Father-specific outcomes, such as paternal help-seeking intention, stigma, paternal self-efficacy, or engagement with professional support pathways, were rarely measured; most studies instead relied on general mental health, relational, or usability measures. In studies where fathers were included within broader couple-based or expectant-parent interventions, paternal outcomes were not always reported separately from maternal or dyadic outcomes. Most studies relied on self-report measures, and follow-up periods were generally short, with no studies reporting outcomes beyond the early postpartum period.

## Discussion

### Principal Findings

This scoping review identified 6 studies examining technology-based, digitally delivered, or hybrid interventions relevant to paternal perinatal mental health. Overall, the evidence is best understood as an emerging intervention landscape rather than a mature effectiveness literature. Several studies focused primarily on feasibility, acceptability, usability, engagement, or intervention design, while others reported early outcome evaluation. This indicates that the field is beginning to move beyond protocol-level intervention development, but it has not yet produced a consistent or robust evidence base from which firm conclusions about effectiveness can be drawn.

A key distinction concerns the role of digital delivery within the intervention model. Because digital tools functioned as the primary support mechanism in some studies and as only a supplementary component in others, the effectiveness of the digital component cannot be interpreted uniformly. In hybrid interventions, observed outcomes may reflect the combined effects of face-to-face, printed, professional, relational, and digital components, rather than the independent contribution of digital delivery itself.

The interventions also differed in design logic, ranging from brief, low-intensity, repeated contact (such as SMS prompts and mood-tracking check-ins) to more structured formats (such as a multiweek course or a multifeature app), with Rover [11] representing the clearest example of a father-specific digital design. These differences suggest that digital technology is not being used in a single consistent way across the literature; rather, it functions variously as a delivery platform, engagement tool, follow-up mechanism, self-monitoring system, or design feature.

Outcome measurement varied considerably across the 6 studies. Mental health symptoms were the most commonly measured domain, with several studies using validated measures such as the EPDS, Kessler Psychological Distress Scale-6, Patient Health Questionnaire-9, Generalized Anxiety Disorder-7, Hospital Anxiety and Depression Scale, or Depression Anxiety Stress Scales-21. This suggests partial convergence around depression and psychological distress measurement, particularly through the EPDS. However, there was no shared outcome framework across studies. Relational and adjustment-related outcomes, including dyadic adjustment, marital quality, psychological resilience, parenting self-efficacy, caregiving quality, and parent-infant bonding, were measured inconsistently and often reflected the specific focus of each intervention. Engagement outcomes were also operationalized differently, including attendance, retention, link clicks, app interactions, mood-tracker responses, lesson adherence, and qualitative feedback. This variation limits comparison across studies and makes it difficult to identify which intervention formats or components are associated with stronger engagement or more meaningful outcomes.

Across the evidence base, findings were more consistent for feasibility, acceptability, usability, and engagement than for clinical effectiveness. SMS delivery, online group discussion, mobile app features, brief digital contact, and digital follow-up were generally reported as acceptable or usable by participants. Features such as father-specific messaging, humor, baby-centered content, mood tracking, mindfulness activities, goal setting, peer normalization, and male facilitation appeared to support engagement in some studies. However, these features were not systematically evaluated across all included studies, and positive acceptability should not be interpreted as evidence of effectiveness. While several studies reported promising engagement or user experience findings, evidence that these interventions produce sustained behavioral change, improved help-seeking, or reduced paternal perinatal depression remains limited.



XSL•FO
RenderX

Effectiveness findings were mixed and should be interpreted cautiously. Some studies reported improvements or protective effects in relational and well-being–related outcomes, such as dyadic adjustment, happiness, psychological resilience, marital quality, or perceived social connection. However, effects on depression, anxiety, parenting stress, and other mental health outcomes were less consistent. In addition, several studies measured paternal mental health indirectly or as a secondary outcome, and follow-up periods were generally short. This limits the extent to which conclusions can be drawn about whether technology-based or hybrid interventions specifically reduce paternal perinatal depression.

Included studies also varied in the extent to which they targeted father-specific outcomes, such as paternal help-seeking, stigma reduction, paternal self-efficacy, parenting confidence, or engagement with professionals, which limits the interpretation of relevance to fathers; a literature gap addressed further in the *Limitations* section below.

## Implications for Digital Intervention Design

The gaps identified in this scoping review point to several priorities for future research. Most pressing is the need for primary studies designed specifically to address paternal perinatal mental health, rather than positioning fathers as secondary participants in mother-focused or general parent-focused interventions. Father-specific interventions should explicitly address fathers' mental health needs, barriers to help-seeking, stigma, paternal identity transition, partner communication, and father-infant bonding.

Future studies should also articulate a clear theoretical basis for intervention design. At present, the included studies vary in whether and how they describe theoretical foundations or mechanisms of change. Clearer theoretical grounding would allow comparison across interventions and help identify which components are most likely to influence outcomes such as mental health literacy, help-seeking intention, emotional regulation, relational support, and engagement with professional care. In addition, future studies should use validated, father-specific outcome measures where possible and have longer follow-up periods to determine whether any benefits are sustained beyond the immediate postpartum period.

A promising avenue for future research includes serious games and gamified interventions. Although 1 included study incorporated gamification elements, no study evaluated a fully developed serious game for paternal perinatal mental health. Game-based approaches may offer novel ways to support mental health literacy, emotional reflection, coping practice, help-seeking attitudes, partner communication, and a means of addressing known barriers to fathers' help-seeking such as time and stigma.

## Limitations of the Included Studies

The included studies have several limitations that affect interpretation of this review. First, although the updated search identified 6 studies, the evidence base remains small and heterogeneous, limiting direct comparison across studies and preventing firm conclusions about the effectiveness, scalability, or suitability of digital interventions for paternal perinatal depression.

Second, not all included studies positioned fathers' postnatal mental health as the primary intervention target. Some studies were designed specifically for fathers, whereas others included fathers or partners within broader couple-based or expectant-parent interventions. While these studies are relevant to paternal postnatal mental health, they do not always isolate paternal outcomes or address fathers' unique barriers to help-seeking, stigma, masculine norms, fatherhood identity, or father-infant bonding. As a result, conclusions about father-specific digital intervention design should be interpreted with caution.

Third, most studies relied heavily on self-report data and short follow-up periods. While self-report measures are appropriate for assessing perceived symptoms, usability, and acceptability, they may be affected by recall bias, social desirability bias, or participant expectations. The limited duration of follow-up also means that little is known about whether digital interventions can produce sustained changes in paternal mental health, help-seeking behavior, relationship functioning, or parenting confidence beyond the early postpartum period.

Fourth, the diversity and representativeness of participants were limited across the included evidence. Several studies used relatively small, educated, partnered, or culturally homogeneous samples. This restricts the generalizability of findings to fathers from diverse cultural, socioeconomic, geographic, linguistic, or family-structure backgrounds. Further research is needed to understand whether digital interventions are acceptable, accessible, and relevant for fathers who may experience additional barriers to support, including fathers in low-resource settings, nonresidential fathers, same-sex parents, and fathers from culturally and linguistically diverse communities.

## Limitations of This Review

Although the search strategy was broad to capture digital, online, mobile, SMS-based, and game-informed interventions, relevant studies may still have been missed due to variation in terminology across digital health, perinatal mental health, and fatherhood research. The review was limited to peer-reviewed English-language publications, which may have excluded relevant gray literature, non-English studies, service evaluations, or unpublished digital health projects.

The search did not include CINAHL or Web of Science in the database searches due to resource constraints, both of which index nursing, allied health, and interdisciplinary literature. Given the overlap between these databases and those searched, it is unlikely that a substantial amount of relevant literature was missed; however, some relevant studies may have been overlooked as a result. Future reviews could include broader database coverage to minimize the risk of omission.

In addition, title screening was conducted by a single reviewer, which may have introduced selection bias, although later stages of screening involved independent review and consensus discussion. Consistent with scoping review methodology, this review aimed to map the available evidence rather than formally appraise study quality or pool effect sizes. Therefore, the



XSL•FO
RenderX

findings should be interpreted as a descriptive and interpretive overview of an emerging evidence base, rather than as evidence of clinical effectiveness.

## Conclusion and Future Work

While the limited evidence base precludes definitive conclusions, the 6 included studies tentatively suggest that digital modalities may offer feasible and acceptable ways of engaging some fathers and partners during the perinatal period. Digital modalities may help reduce father-specific barriers related to time, geography, stigma, and reluctance to attend face-to-face services; however, future research is required to systematically evaluate and confirm these potential advantages.

Intervention features such as father-specific messaging, humor, baby-centered content, brief digital contact, mood tracking, goal setting, mindfulness activities, and peer normalization were described as potentially supportive of engagement. However, the evidence that these features translate into sustained behavior change, improved help-seeking, or reduced paternal PND remains limited. Acceptability and usability findings should therefore be understood as early design evidence that still requires proof of clinical effectiveness.

The current evidence base remains too small and heterogeneous to support firm conclusions about the scalability, suitability, or effectiveness of digital interventions for fathers experiencing perinatal mental health challenges. Future research should prioritize father-specific digital interventions that are theory-informed, co-designed with fathers, and evaluated using rigorous methods, including validated father-specific outcomes, longer follow-up periods, transparent reporting of engagement and attrition, and more diverse participant samples.

Fathers remain a consistently underserved population despite the prevalence of paternal perinatal mental health challenges. Digital interventions offer a promising and timely avenue for future research, with the potential for innovative, father-tailored solutions that could meaningfully advance both clinical practice and public health initiatives in this critical area.

## Acknowledgments

The authors acknowledge the use of generative AI (GenAI), namely Grammarly's "Superhuman Platform" and OpenAI's GPT-5 in the writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: text generation, proofreading and editing, reformatting, and publication support. All critical interpretation, analysis, conceptualization, and conclusions presented in this work were conducted by the authors and responsibility for the final manuscript lies entirely with the authors.

## Funding

The authors declared no financial support was received for this work.

## Authors' Contributions

Conceptualization: WW, EP, JG
Data curation: WW
Formal analysis: WW
Investigation: WW, EP
Methodology: WW, EP, JG
Supervision: EP
Writing – original draft: WW
Writing – review and editing: EP, JG

## Conflicts of Interest

None declared.

Checklist 1
PRISMA-ScR checklist.
[PDF File, 563 KB - i-jmr_v15i1e95238_app1.pdf ]

## References

1. Rodrigues AL, Ericksen J, Watson B, Gemmill AW, Milgrom J. Interventions for perinatal depression and anxiety in fathers: a mini-review. Front Psychol 2021;12:744921. [doi: 10.3389/fpsyg.2021.744921] [Medline: 35126228]
2. Searson C, Oxlad M, Sawyer A. "Nobody asked me, as a dad, how I was coping": exploring Australian new fathers' paternal postnatal support experiences and needs. Clin Psychol 2025 May 4;29(2):131-144. [doi: 10.1080/13284207.2025.2494512]
3. Hodgson S, Painter J, Kilby L, Hirst J. The experiences of first-time fathers in perinatal services: present but invisible. Healthcare (Basel) 2021 Feb 3;9(2):161. [doi: 10.3390/healthcare9020161] [Medline: 33546202]



XSL•FO
RenderX

4. Ghaleiha A, Barber C, Tamatea AJ, Bird A. Fathers' help seeking behavior and attitudes during their transition to parenthood. Infant Ment Health J 2022 Sep;43(5):756-768. [doi: 10.1002/imhj.22008] [Medline: 35913697]
5. Darwin Z, Galdas P, Hinchliff S, et al. Fathers' views and experiences of their own mental health during pregnancy and the first postnatal year: a qualitative interview study of men participating in the UK Born and Bred in Yorkshire (BaBY) cohort. BMC Pregnancy Childbirth 2017 Jan 26;17(1):45. [doi: 10.1186/s12884-017-1229-4] [Medline: 28125983]
6. Cameron EE, Sedov ID, Tomfohr-Madsen LM. Prevalence of paternal depression in pregnancy and the postpartum: an updated meta-analysis. J Affect Disord 2016 Dec;206:189-203. [doi: 10.1016/j.jad.2016.07.044] [Medline: 27475890]
7. Pilkington PD, Whelan TA, Milne LC. A review of partner‐inclusive interventions for preventing postnatal depression and anxiety. Clin Psychol 2015 Jul 1;19(2):63-75. [doi: 10.1111/cp.12054]
8. Rominov H, Pilkington PD, Giallo R, Whelan TA. A systematic review of interventions targeting paternal mental health in the perinatal period. Infant Ment Health J 2016 May;37(3):289-301. [doi: 10.1002/imhj.21560] [Medline: 27079685]
9. Cameron EE, Hunter D, Sedov ID, Tomfohr-Madsen LM. What do dads want? Treatment preferences for paternal postpartum depression. J Affect Disord 2017 Jun;215:62-70. [doi: 10.1016/j.jad.2017.03.031] [Medline: 28319693]
10. Goldstein Z, Rosen B, Howlett A, Anderson M, Herman D. Interventions for paternal perinatal depression: a systematic review. J Affect Disord 2020 Mar 15;265:505-510. [doi: 10.1016/j.jad.2019.12.029] [Medline: 32090778]
11. Teague SJ, Shatte ABR, Fuller-Tyszkiewicz M, Hutchinson DM. User-centred design and evaluation of an mHealth app for fathers' perinatal mental health: a feasibility, acceptability, and usability study. Behav Inf Technol 2025;44(20):4950-4964. [doi: 10.1080/0144929X.2025.2502474] [Medline: 40385514]
12. Xie EB, Jung JW, Kaur J, Benzies KM, Tomfohr-Madsen L, Keys E. Digital parenting interventions for fathers of infants from conception to the age of 12 months: systematic review of mixed methods studies. J Med Internet Res 2023 Jul 26;25:e43219. [doi: 10.2196/43219] [Medline: 37494086]
13. de-Juan-Iglesias P, Gómez-Gómez I, Barquero-Jimenez C, Wilson CA, Motrico E. Effectiveness of online psychological interventions to prevent perinatal depression in fathers and non-birthing partners: a systematic review and meta-analysis of randomized controlled trials. Internet Interv 2024 Sep;37:100759. [doi: 10.1016/j.invent.2024.100759] [Medline: 39149135]
14. Perinatal mental health screening in Australia. : AIHW (Australian Institute of Health and Welfare), Australian Government; 2024 URL: https://www.aihw.gov.au/getmedia/5ca67080-f588-422d-9115-04e96b0b36fa/perinatal-mental-health-screening-in-australia.pdf?v=20260116145231&inline=true [accessed 2026-08-11]
15. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310] [Medline: 843571]
17. Fletcher R, Kay-Lambkin F, May C, Oldmeadow C, Attia J, Leigh L. Supporting men through their transition to fatherhood with messages delivered to their smartphones: a feasibility study of SMS4dads. BMC Public Health 2017 Dec 13;17(1):953. [doi: 10.1186/s12889-017-4978-0] [Medline: 29237439]
18. Missler M, van Straten A, Denissen J, Donker T, Beijers R. Effectiveness of a psycho-educational intervention for expecting parents to prevent postpartum parenting stress, depression and anxiety: a randomized controlled trial. BMC Pregnancy Childbirth 2020 Oct 31;20(1):658. [doi: 10.1186/s12884-020-03341-9] [Medline: 33129314]
19. Canfield SM, Canada KE, Rolbiecki AJ, Petroski GF. Feasibility and acceptability of an online mental health intervention for pregnant women and their partners: a mixed method study with a pilot randomized control trial. BMC Pregnancy Childbirth 2023 Oct 18;23(1):739. [doi: 10.1186/s12884-023-06031-4] [Medline: 37853333]
20. Karakoç H, Kalkan E. The effect of paternal support on dyadic adjustment, happiness and psychological resilience: a randomized controlled trial. Curr Psychol 2024 Nov;43(42):32921-32930. [doi: 10.1007/s12144-024-06833-5]
21. Kaner A, Cwikel J, Segal-Engelchin D. The transition to fatherhood—evaluation of an online intervention for new fathers. Psychol Health Med 2024 Jun;29(5):1011-1019. [doi: 10.1080/13548506.2023.2260600] [Medline: 37735804]


## Abbreviations

**CBT:** cognitive behavioral therapy
**EPDS:** Edinburgh Postnatal Depression Scale
**mHealth:** mobile health
**PND:** postnatal depression
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
**RQ:** research question



XSL•FO
RenderX

*Edited by M Balcarras; submitted 12.Mar.2026; peer-reviewed by I Espinosa-curiel, L Ilodigwe, M Evans; revised version received 26.Jul.2026; accepted 27.Jul.2026; published 26.Aug.2026.*

*Please cite as:*
*Wang W, Penney E, Garcia J*
*Health Information Technologies in the Support System of New Fathers' Postnatal Depression: Scoping Review*
*Interact J Med Res 2026;15:e95238*
*URL: https://www.i-jmr.org/2026/1/e95238*
*doi:10.2196/95238*

© Weijie Wang, Erika Penney, Jaime Garcia. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 26.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Contexts and Mechanisms Related to the Efficacy of Digital Medication Adherence Interventions to Support Medication Adherence Among Adults With Chronic Diseases: Realist Review

Srinjaya Saha[1*], PhD; Yvonne Latham[2*], PhD; Carol Holland[3*], PhD

[1]Department of Health Research, Faculty of Health and Medicine, Lancaster University, Health Innovation Campus, John Fisher Drive, Bailrigg, Lancaster, England, United Kingdom

[2]Department of Organisation, Work and Technology, Lancaster University Management School, Lancaster University, Bailrigg, Lancaster, England, United Kingdom

[3]Department of Health Research, Faculty of Health and Medicine, Lancaster University, Bailrigg, Lancaster, England, United Kingdom

[*]all authors contributed equally

**Corresponding Author:**
Srinjaya Saha, PhD
Department of Health Research, Faculty of Health and Medicine, Lancaster University, Health Innovation Campus, John Fisher Drive, Bailrigg, Lancaster, England, United Kingdom

## Abstract

**Background:** Digital medication adherence interventions for individuals diagnosed with chronic conditions have been developed, but their long-term efficacy in improving medication adherence has been limited. Exploring mechanisms and contexts related to outcomes of digital medication adherence interventions is important for these to be effectively tailored for different populations and contexts. A realist review was conducted to assess which intervention components might work for whom, and under what circumstances.

**Methods:** A realist review of literature published between 2002 - 2024 was conducted. Studies reporting digital medication adherence interventions for people diagnosed with chronic diseases and experiencing unintentional medication nonadherence issues were included. MEDLINE, CINAHL, PsycInfo, Web of Science, Scopus, Embase, and gray literature databases, Overton, and Policy Commons were searched. Data regarding contexts, mechanisms, and outcomes were extracted and synthesized into program theories. Program theories from the realist review were triangulated with findings from two rounds of workshops (three workshops in each round), in which context-mechanism-outcome (CMO) configurations were cocreated with various stakeholders of a project aimed at implementing a digital medication adherence intervention to reduce unintentional nonadherence among people with chronic diseases. There were 12 participants in each round.

**Results:** The analysis of 83 papers and content of the cocreation workshops led to 35 CMOs theories in seven theory areas, which were (1) building and maintaining medication intake habits; (2) solving medication-related barriers to adherence; (3) enabling collection of prescription medications on time; (4) supporting people with unpredictable health conditions and sensory or motor impairments; (5) supporting individuals with high anxiety, low medication self-efficacy, and low social support; (6) supporting individuals who are uncomfortable with technology; and (7) economic, policy, and organizational factors affecting implementation of smart medication devices. Fourteen CMOs came from both the literature and workshops, 7 CMOs from the literature alone, and 14 CMOs from the workshops alone.

**Conclusions:** Analysis of the program theories suggested a range of intervention components considering contextual factors that may improve medication adherence of individuals. These were used to form recommendations for intervention developers and implementers. Further studies are required on policy and economic factors affecting large-scale implementation of digital medication adherence interventions in different settings. Future interventions should report intervention content and delivery in detail and codevelop effective implementation strategies with professionals who can support the implementation of the intervention.

**Trial Registration:** PROSPERO CRD42025641383; https://www.crd.york.ac.uk/PROSPERO/view/CRD42025641383

***(Interact J Med Res 2026;15:e89803)*** doi:10.2196/89803

**KEYWORDS**

medication adherence; unintentional nonadherence; digital interventions; realist review; program theories



XSL•FO
RenderX

## Introduction

Medication adherence is defined as the extent to which individuals' medication intake is consistent with their health care providers' recommendations [1]. Medication nonadherence is particularly common among individuals diagnosed with long-term health conditions. Consequences of medication nonadherence include worsening of disease symptoms, increased rate of hospitalization, and mortality [2,3].

Medication adherence is a complex phenomenon, and nonadherence occurs in a range of different scenarios: if individuals do not start taking their prescribed medications (primary nonadherence) to begin with; take wrong doses at the wrong time; skip doses (nonconforming); or stop taking the medications against health care providers' recommendations (nonpersistence) [4,5]. Medication nonadherence could be intentional, whereby an individual actively decides not to take medications as prescribed due to factors such as severe side-effects, or being nonintentional, which is an unplanned behavior possibly because of forgetfulness [1]. The barriers to medication adherence fall under 5 dimensions [2]: social and economic factors (eg, low socioeconomic status); health system/health care related factors (eg, patient-provider relationship); therapy related factors (eg, complexity of medication regimen); condition-related factors (eg, severity of symptoms and comorbidities); and patient related factors (eg, patient's beliefs about perceived need for medications). It has been recommended that personalized interventions that target patients' barriers to medication adherence in all 5 dimensions can help improve medication adherence [2].

However, the findings of a systematic review by Yap et al [6] highlighted the conflicts in evidence regarding the factors associated with medication adherence and showed that barriers to medication adherence could vary among different contexts. This highlights the importance of understanding how contextual factors may influence adherence and designing tailored interventions for specific contexts [7,8].

Digital interventions have been developed to improve medication adherence, such as providing text message reminders to take medications [8,9]. Digital interventions have the potential to contribute to more equitable delivery of care, helping improve health care for patients living in remote or rural areas and facilitating more frequent communication with patients; for example, medication reminders can be sent more frequently through digital interventions [10]. However, systematic reviews have highlighted the lack of evidence of long-term efficacy of digital medication adherence interventions in improving medication adherence among individuals with chronic illnesses. Interventions tailored to individual patients, targeting multiple dimensions of medication nonadherence [1,8] as recommended by the World Health Organization [2], are also lacking.

Systematic reviews have also highlighted the difficulties in understanding the mechanisms by which the interventions may have led to improvement in medication adherence, as the studies included were heterogeneous in terms of populations and content and delivery of the interventions [10,11]. Additionally, systematic and scoping reviews provide information about the efficacy or inefficacy of interventions but rarely provide a detailed analysis of contexts and mechanisms related to intervention efficacy or inefficacy [12]. To advance understanding, we need to study such contexts and mechanisms related to outcomes of digital medication adherence interventions.

To enable this understanding, we used realist review methods. A realist review aims to understand the mechanisms of *how* complex interventions work or do not work in different contexts as opposed to just *whether* they work or not [13]. Contexts (circumstances or conditions in which mechanisms can work and lead to the outcomes), mechanisms (elements of the interventions which can potentially lead to the outcomes), and outcomes are presented together as program theories to explain how interventions work [13]. As opposed to systematic reviews, realist reviews use a wider range of evidence, including gray literature, to understand what works for whom and under what conditions [12,13]. Thus, this study aimed to explore the contexts and mechanisms related to the efficacy or inefficacy of digital medication adherence interventions in improving medication adherence among patients with chronic illnesses experiencing unintentional medication adherence issues. To the best of the authors' knowledge, this is the first realist review to assess contexts and mechanisms related to outcomes of digital medication adherence interventions and provide recommendations for designing and implementing similar digital medication interventions.

## Methods

### Overview

This study was preregistered on PROSPERO (International Prospective Register of Systematic Reviews; CRD42025641383) and has been reported according to the RAMESES (Realist and Meta-Narrative Evidence Syntheses: Evolving Standards) [14].

The glossary of terms used to describe the steps undertaken to carry out this review has been presented in Table 1.



XSL•FO
RenderX

**Table .** Glossary of terms used to describe the process of this realist review.

| Terms | Description |
|---|---|
| Contexts | Conditions in which the intervention operates, including factors related to individuals receiving the intervention and social, cultural, economic, or policy level factors [13,15]. |
| Mechanisms | The underlying processes that make the intervention work, which operate in certain contexts/situations to lead to outcomes. These can be classified as "mechanism resources," the resources provided by the interventions, and "mechanism responses," the responses that the intervention resources elicit among participants of the intervention [16]. |
| Outcomes | The intended and/or unintended results of an intervention [14]. |
| CMO[a] configurations | A CMO configuration is a heuristic used in realist research to show patterns of how an intervention generates outcomes and the mechanisms that operate in contexts associated with them [14]. |
| Program | Interventions or services under investigation [17]. In this realist review, the program refers to the digital interventions that aimed to improve medication adherence or reduce unintentional nonadherence. |
| Program theories | Explanations about how specific interventions work. CMOs represent individual program theories. This realist review also presents overarching program theories, which refer to the integration of individual CMO configurations into explanatory models. |
| Initial program theories | Initial ideas regarding the program theories of an intervention. These CMO configurations are further refined during the realist review process [15]. |

[a]CMO: context-mechanism-outcome.

The following 5 steps were taken to conduct the realist review.

## Development of Initial Program Theory

Initial program theories were cocreated in workshops with the stakeholders of a pilot project that aimed to explore the efficacy of a smart medication device called the "Medybox" in improving medication adherence among domiciliary care recipients and hospital discharge patients in a region in Northern England, United Kingdom. In the workshops, the stakeholders hypothetically discussed the features of the Medybox (a modified electronic pillbox) as an example of a digital medication adherence intervention, which informed the realist review that explored the program theories of all types of digital medication adherence interventions to reduce unintentional medication nonadherence. These stakeholders were presented with a summary of the features of the device and its implementation process, summarized as follows:

- Potential recipients of this device were adults diagnosed with any chronic disease who were unintentionally nonadherent to their medication regimens.
- Domiciliary care providers, short-term re-enablement services, and hospital discharge teams were to identify individuals who needed support to improve their medication adherence and refer them to use the device.
- The Medybox device was a modified electronic pillbox in that the device did not store medications according to individual doses. Instead, it safely stored all types of medications (except injectables and medications that needed refrigeration), including larger liquid medications. The device provided light and sound reminders for medication intake and transmitted real-time responses to the reminders (opening of the device by the participants), allowing the device providers to remotely monitor individuals' medication adherence and provide feedback on it. The device had a built-in SIM card and did not require a Wi-Fi connection. It was to be provided to individuals for 6 - 12 weeks, and during this time the device was intended to help individuals form effective habits of taking their medicines correctly and independently, after which it would be taken away. After that, the device providers planned to recommend other types of interventions or technologies to maintain recipients' medication adherence, such as mobile apps, with a return to caregiver support as one option.

Six cocreation workshops of approximately 2 hours each in 2 workshop rounds (3 workshops and 12 participants in each round) were conducted to generate ideas for program theories for the implementation of this particular device. Stakeholders included members of the regional National Health Service integrated care board (ICB) who were former health care professionals (nurses) but currently working on a digitizing adult social care program, device providers who were implementing the intervention, ICB pharmacists, managers of domiciliary care organizations, individuals receiving care for their medication intake and their informal or family caregivers. Most of the stakeholders were women (10/12, 83%) and 83% (10/12) were White British (consistent with the ethnicity of the population in the local area).

In keeping with qualitative methods, a small number of appropriate stakeholders were recruited purposively via word of mouth and contacts of the research team, the ICB, and the device providers. Stakeholders were "experts by experience" as they had lived experiences of receiving or providing care to individuals requiring help with medication adherence. ICB professionals were also knowledgeable about possible contextual factors that could lead to unintended outcomes of the program



XSL•FO
RenderX

[13,18]. The same group of professionals and care recipients were invited to participate in the second round of workshops. Additionally, participants were recruited if a certain group of stakeholders was underrepresented in the first round. One of the device providers involved in the implementation of the program dropped out after the first round. Therefore, the researchers recruited an ICB pharmacist in the second round to maintain numbers and improve diversity of the stakeholders involved in the workshops (see Saha et al [19] for detailed information about cocreation workshops). The workshops did not include academic researchers as stakeholders.

The discussions were facilitated by the research team, who provided participants with prompts regarding potential contexts (eg, technological literacy of individuals), mechanisms (medication intake reminders), and outcomes (eg, medication adherence) that were derived from the emerging findings of this review. Specific parts of the workshop discussions were audio- or audio-and-video-recorded, and the research team made notes of the workshop discussions. The context-mechanism-outcomes (CMOs) derived from the first round of workshops were presented to the participants in the second round of workshops to further refine them. A list of initial ideas of CMOs derived from workshop discussions was triangulated with the findings from the realist review. This was an iterative process in which the workshop discussions informed the realist review and vice versa. Further details about the device and the prompts used to facilitate workshop discussions are presented in Multimedia Appendix 1. This method of conducting cocreation workshops has been used to explore initial program theories by other realist reviews [20].

## Literature Search

Searches were conducted on the following databases from January 2002 till December 2024: MEDLINE, CINAHL, PsycInfo, Web of Science, Scopus, Embase, and gray literature databases called Overton and Policy Commons. Hand searching of reference lists of reviews was conducted to identify any additional studies. The research team started searching from January 2002, as technologies such as smart pillboxes (devices that store and dispense scheduled medication doses at correct times, provide medication intake reminders, and monitor medication adherence by recording when the user opens the pillbox), for example, Pivotell, were first used extensively in the United Kingdom in that year [21]. Thus, these search dates increased the likelihood of inclusion of studies related to recent digital technologies designed to improve medication adherence such as smart pillboxes, automated interactive voice recognition, and mobile apps [22]. The search terms were decided based on the population (eg, multiple chronic conditions), intervention (eg, digital intervention), and outcomes (eg, medication adherence) explored in this review. The full search terms are presented in Multimedia Appendix 2.

## Screening and Selection

The papers were assessed for eligibility based on the inclusion and exclusion criteria presented in Table 2.

Titles and abstracts were screened by the first author using the web-based systematic review platform Rayyan [23] and around 10% (200/2152) of the papers were also screened by the other members of the research team for inclusion. Any disagreements were resolved through discussion.



XSL•FO
RenderX

**Table .** Inclusion and exclusion criteria used for selecting studies in the review.

| | Inclusion | Exclusion |
|---|---|---|
| Population | Adults diagnosed with a chronic disease and taking long-term medicines to manage it. | • Individuals receiving medication adherence support in nursing homes and in-patient settings such as hospitals and prisons, as the medications are managed by professionals in these contexts [24,25].<br>• Individuals younger than the age of 18 years.<br>• Individuals having fluctuating capacity to consent, such as those diagnosed with dementia or substance abuse disorders.<br>• Studies that focus on one specific illness diagnosis, as the CMOs[a] from the review were triangulated with those from the cocreation workshops, and the workshop discussions were centered around a device aimed at improving adherence for people diagnosed with any type of chronic illness. Additionally, papers focusing on only one type of chronic illness are unlikely to provide detailed information for program theory development as the information reported would be relevant to specific chronic illness diagnoses only. |
| Intervention | Studies reporting an intervention to improve medication adherence or reduce unintentional nonadherence with at least one digital component such as websites, mobile apps, or other digital devices. No restrictions were applied on the type of digital adherence intervention to reduce unintentional nonadherence. | • Digital medication adherence interventions that only include content designed to reduce intentional nonadherence, such as education about the importance of taking prescribed medications correctly. This was based on previous studies that have consistently shown that multicomponent interventions are more effective in improving adherence as compared to interventions involving patient education only [26,27]. This criterion was also used because the workshop discussions (added qualitative data) were aimed at developing CMOs of an example device aimed at reducing unintentional nonadherence. |
| Outcomes | Studies assessing changes in medication adherence or unintentional nonadherence and/or contexts and mechanisms and outcomes of the interventions. | • Did not report medication adherence and/or unintentional nonadherence and/or their related contexts, mechanisms, and outcomes. |
| Document types | Consistent with the guidelines for conducting realist reviews [14,28], no restrictions were placed on study designs. Reviews, opinion papers, and gray literature reports were included if they provided sufficient information about contexts, mechanisms, and outcomes. | • Conference abstracts, paper corrections, intervention protocols, retractions, and book reviews as they are unlikely to have sufficient information for program theory development. |

[a]CMO: context-mechanism-outcome.

## Data Extraction

A data extraction sheet was used to document details regarding authors, year of publication, study title, document type (eg, systematic review and pre-post intervention study), participants, and rigor, richness, and relevance scores. The data extraction sheet is presented in Multimedia Appendix 3 [3,9-11,29-107]. The rigor of the studies was judged based on the credibility and trustworthiness of the methodology, for example, sample size, data collection, and analysis methods [13]. The rigor of gray literature, opinion pieces, and book chapters was assessed based on significance, objectivity, coverage of the topic, authority, and accuracy [28]. Relevance was judged by assessing the extent to which papers presented information related to the topic of the review and provided evidence for program theory development. Richness was judged by assessing the extent to which papers provided details regarding mechanisms, contexts, and outcomes of the interventions [12,13]. The scoring criteria used in the realist review by Waldron et al [12] were used to assess the rigor, relevance, and richness. Studies were scored on a scale of 0 (very poor) to 3 (very good) in terms of relevance



XSL•FO
RenderX

and rigor, and 0 (nothing of interest or not focused on design, implementation, or use) to 4 (much valuable data) in terms of richness. This helped in understanding the extent to which each study contributed to program theory development and the extent to which each program theory was supported by relevant and rigorous evidence [12,13].

### Data Synthesis

The lead author manually annotated the papers for program theory ideas with the help of NVIVO (Lumivero, version 14) software. Data from the literature and workshops were analyzed separately. Patterns of context, mechanisms, and outcomes from the studies included were explored by using inductive and deductive logic, and relevant text to support each program theory was documented. The recordings of the workshop discussions were transcribed verbatim. The lead author reviewed the transcripts and the participant and facilitator notes of the workshop discussions and coded the patterns of contexts, mechanisms, and outcomes. Anonymized participant quotes and participant and facilitator notes supporting each CMO were documented.

The CMOs from the realist review were then triangulated with the CMOs from the cocreation workshops by noting the extent to which the CMOs from the workshops corroborated or not with the CMOs derived from the literature. After that, a final list of program theories in different theory areas and the number of CMOs from the literature, workshops, or both were noted. After the initial CMOs were created, an iterative process of revising and refining was conducted through discussion with the research team.

### Ethical Considerations

This study received ethics approval from the Clinical Research Sponsorship ethics committee at Lancaster University and the Health Research Authority and Health and Care Research Wales ethics committee on February 4, 2025 (ethics board: social care; reference number: 24/IEC08/0037).

## Results

### Characteristics and Quality of Studies

Eighty-three studies were included in this review. Figure 1 depicts the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram with details about the search and screening process.

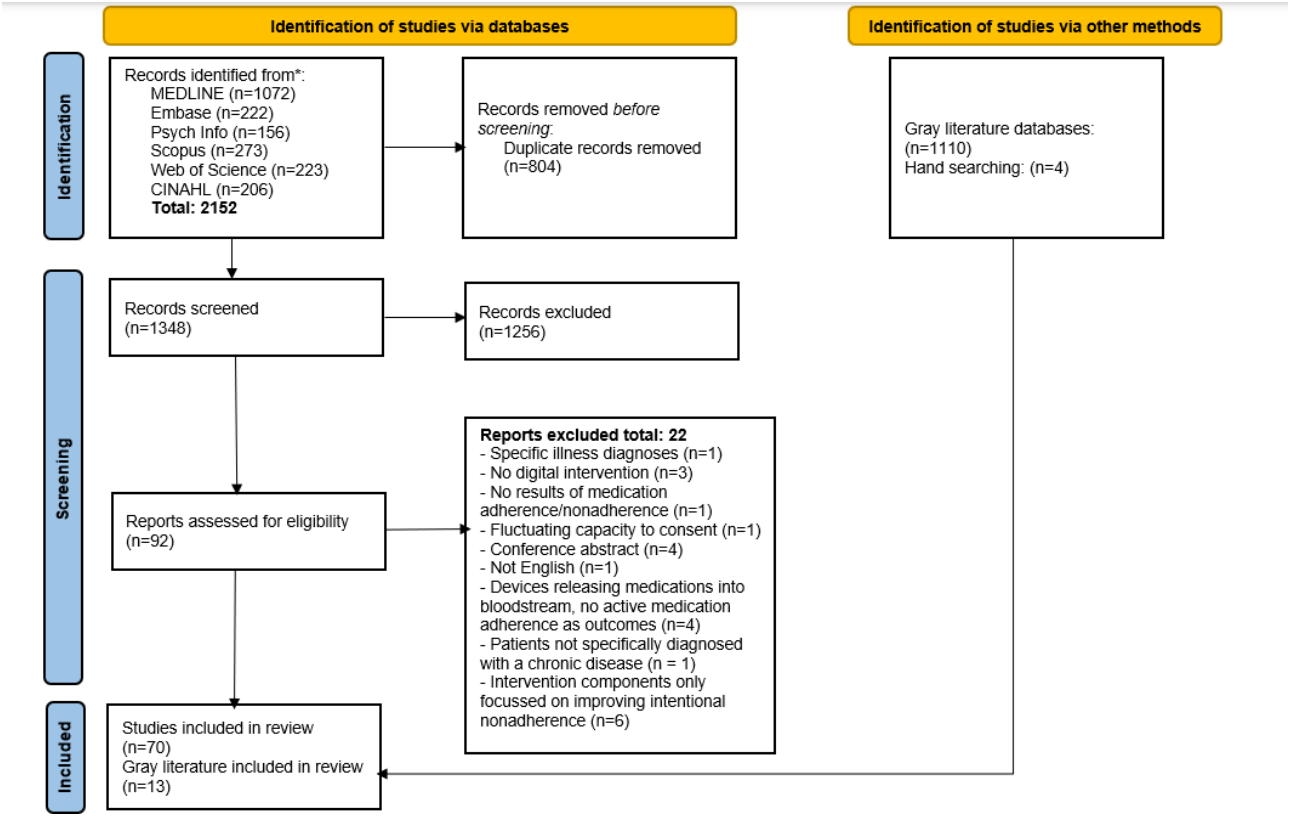
Eighty-three papers, including peer-reviewed research papers (n=70) and gray literature (n=13), were included in the review. Of the research papers, 44 were narrative, systematic, or scoping reviews or meta-analyses, and 3 papers were expert opinion pieces. Of the 23 primary studies included, the largest number were conducted in the United States (n=9), with studies also conducted in Canada (n=4), Spain (n=2), the United Kingdom (n=2), China (n=2), Australia (n=1), and Switzerland (n=1). The country in which the study was conducted was not reported by 2 studies [29,30]. The primary studies included 17 intervention studies [3,10,29-44]. The rest were qualitative or quantitative studies that explored participants' views (where participants were individuals diagnosed with chronic conditions, informal caregivers of patients diagnosed with chronic conditions, and health care professionals) regarding the components of digital medication adherence interventions [43-48].

Most of the studies were assigned a score of 3 in the "rigor" domain. However, approximately 60% (50/83) of the studies were assigned a score of 1 or 2 in the richness domain due to lack of detail regarding contexts and mechanisms related to the outcomes of the interventions. This was primarily because these studies did not report the content and delivery of the interventions in sufficient detail. Approximately 24% (20/83) of the studies were assigned a score of 1 in the "relevance" domain due to lack of data for program theory development. The authors triangulated data from studies that had low "rigor" scores with the data from studies that had higher scores while developing program theories [13,108]. The studies that had low "relevance" and "richness" scores were included in the review if they provided enough information for program theory development and if the data provided by them could be supported by the studies that were assigned high scores in these domains. The authors also transparently reported if a certain program theory was not supported adequately by rigorous, relevant, and rich data [108]. Details about the studies included in the review are presented in Multimedia Appendix 3.



XSL•FO
RenderX

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram [109] with details regarding the inclusion of studies.




Thirty-five CMOs grouped in 7 theory areas were derived from the review, cocreation workshops, or both. A narrative summary of the CMOs derived from the literature, cocreation workshops, or both in each theory area is presented in the following section. The CMOs are presented in Multimedia Appendix 4 [32,35,38,40-42,45,47,48,52,56,74,79,87,97,100,106], along with specific quotes to support them that were derived from workshop discussion recordings and/or studies in this review.

## Theory 1: Building and Maintaining Medication Intake Habits

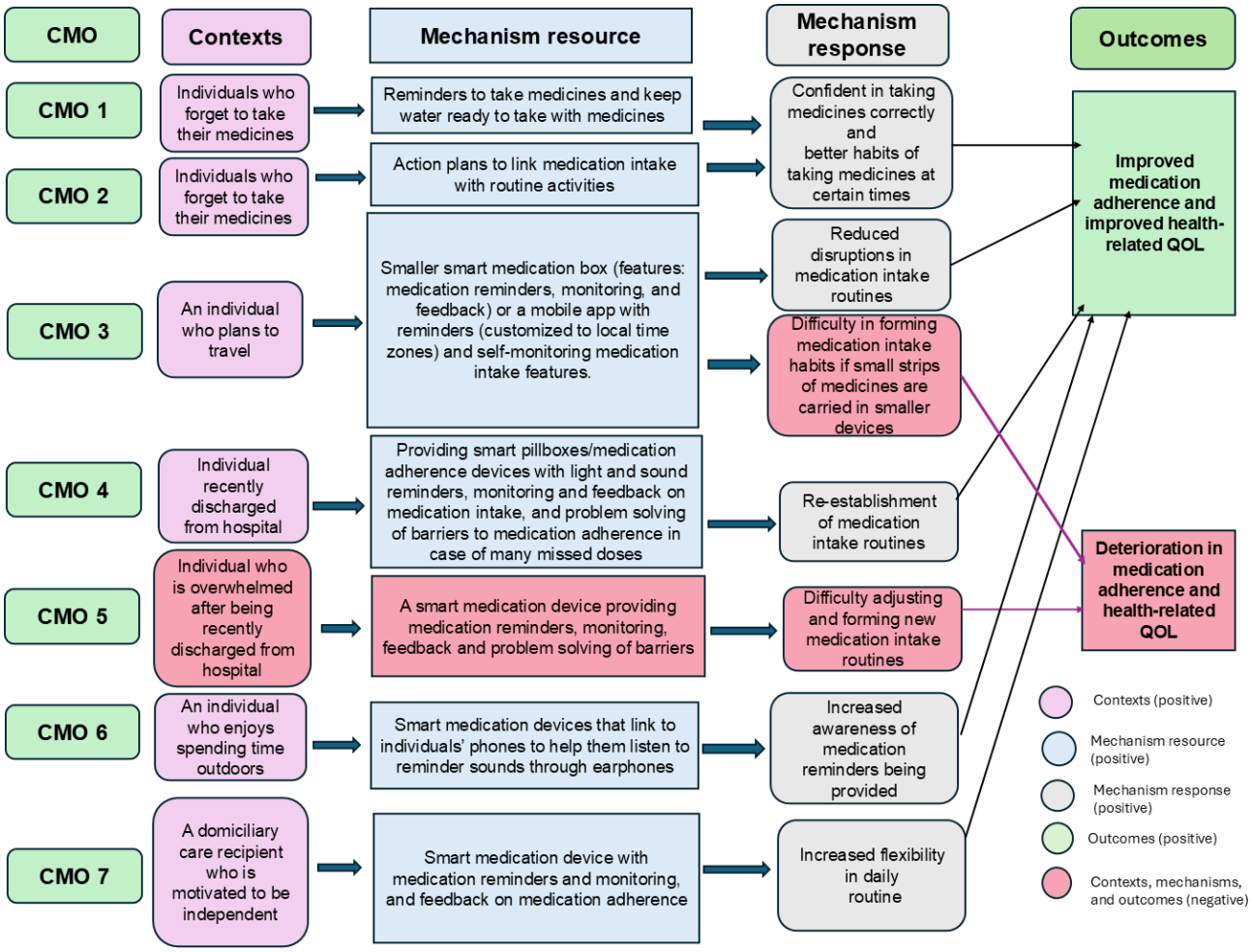
### *Overview*

This theory area includes the CMOs related to ways by which digital interventions facilitate the development and maintenance of effective medication intake habits. The overarching program theory based on the CMOs of theory 1 is presented in Figure 2 (see also Multimedia Appendix 4).



XSL•FO
RenderX

**Figure 2.** The overarching figure showing the CMOs in theory 1: building and maintaining medication intake habits. CMO: context-mechanism-outcome; QOL: quality of life.




### *CMOs Derived From Both Literature and Workshops*

Providing medication intake reminders using various devices, including mobile apps, text messages, or smart medication dispensers, can help individuals to form effective medication intake habits [10,31-46,49-97]. Additionally, workshop participants suggested that it was important to prompt individuals to keep water ready and near the smart medication device, as this can help individuals to establish habits of undertaking the necessary actions before medication intake. However, it is important to tailor the frequency and timing of reminders according to individual needs. Providing reminders at the exact times when patients plan to take their medicines was recommended to avoid excessive intrusiveness in individuals' lives (CMO 1) [58].

Monitoring individuals' medication adherence using smart medication devices or mobile apps and collaboratively developing action plans with them to link their medication intake with daily routine tasks, such as mealtimes [9,30,39,87,95], can also help in forming effective medication intake habits. Over time, these tasks could act as cues for medication intake, resulting in lower chances of forgetfulness and improved adherence (CMO 2).

Certain situations could disrupt existing habits related to medication intake. Hospitalization was identified as one such situation, and it could be difficult to form new and effective medication intake habits after hospital discharge, as hospitalization commonly results in medication regimen changes [35]. Thus, providing patients with smart medication devices that provide medication intake reminders and allow monitoring and feedback on medication adherence can lead to the re-establishment of effective medication intake habits and improvement in adherence (CMO 3 [35]). However, some individuals may be overwhelmed by the recent hospitalization and may not want any more changes to their routines after discharge, resulting in refusal to use smart medication devices. This could be a risk if the individual does not have any informal caregivers or care providers to help them take their medications correctly, which can potentially result in poor medication adherence and rehospitalization (CMO 4) [35].

Traveling could also lead to disruption in medication intake habits. The workshop discussions centered around a device big enough to effectively store all the prescribed medications, including liquids, but too difficult to carry while traveling. Therefore, it was suggested that the device should be left at home while traveling and individuals could use a mobile app to receive medication intake reminders and self-monitor their medication intake. Additionally, a qualitative study highlighted that mobile apps should provide medication intake reminders that are customizable to match local time zones when the individual is traveling [40]. The use of smaller smart medication



XSL•FO
RenderX

devices while traveling was also suggested by a few workshop participants. However, this could lower medication adherence if an individual takes out small strips of medications to carry with them, as they would not have access to instructions regarding medication intake that are printed on medication boxes (CMO 5).

### *CMOs Derived Only From Workshops*

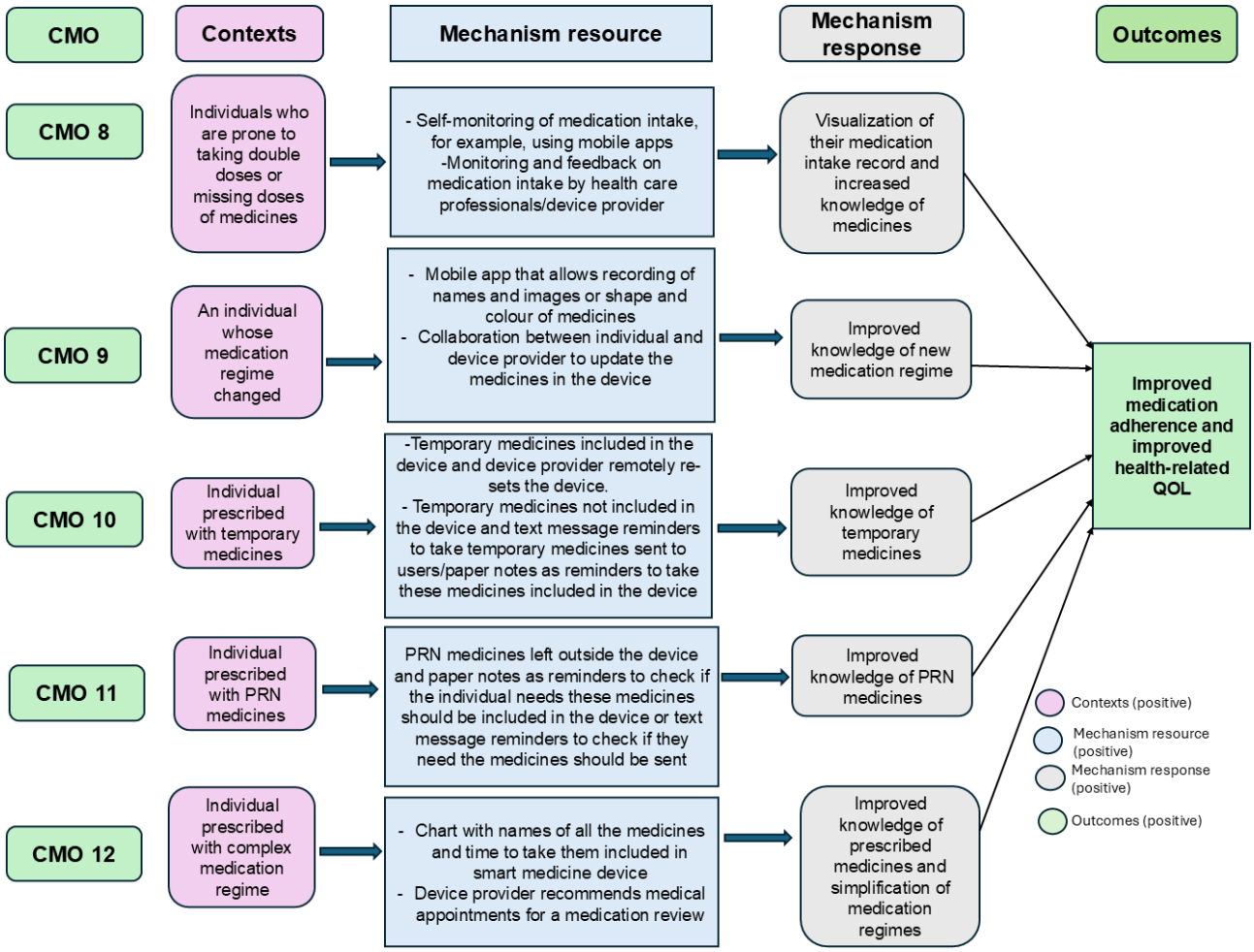
Domiciliary care providers often visit domiciliary care recipients to prompt them to take their medications on time. Therefore, workshop participants mentioned that this device could help care recipients be more independent and flexible as they do not have to wait for care providers to visit them to take their medications (CMO 6). Workshop participants also discussed how medication reminder sounds provided by the device may not be heard by some individuals if they spend a lot of time outdoors, for example, in their garden. Thus, it was suggested that the device could be connected to individuals' mobile phones so that they could hear the reminder sounds through earbuds while engaging in activities outdoors (CMO 7).

## Theory 2: Solving Medication-Related Barriers to Adherence

### *Overview*

This theory area includes the CMOs related to ways by which digital interventions help in solving medication-related barriers to adherence, such as complex regimens. The overarching program theory presented in Figure 3 is based on the CMOs of theory 2 presented in Multimedia Appendix 4.

**Figure 3.** The overarching figure showing the CMOs in theory 2: solving medication-related barriers to adherence. CMO: context-mechanism-outcome; QOL: quality of life; PRN: pro re nata.




### *CMOs Derived From Both Literature and Workshops*

Some individuals may be more prone to frequently missing doses of medications or taking double doses due to factors such as forgetfulness. Digital medication adherence interventions such as mobile apps can help in improving adherence by allowing individuals to self-monitor their medication intake. This can reduce chances of double doses or missed doses of medications due to individuals' increased awareness of their own medication intake [3,11,45,46,48,52,54-56,63,73,88,96,98]. Smart medication devices, which allow a health care professional or provider of the smart medication device to monitor and provide feedback on an individual's medication adherence, can also help in reducing chances of missed doses and double doses, as they can increase individuals' knowledge of their medications (CMO 8) [9,29,30,39,48,52,59,63,79,85,86,89,93,96,98].

Individuals' medication adherence can also be reduced because of medication regimen changes, as new medication intake habits may need to be established. In this context, mobile apps that enable recording of medication images or that include an option



XSL•FO
RenderX

of adding the shape and color of medications along with names of the medications can help improve the recognition of new prescribed medications [32,40]. Mobile apps that announce medication names and show the pictures of the medications while providing medication intake reminders can also help in improving medication adherence by improving recognition and knowledge of the medications [40]. Additionally, workshop participants suggested that there should be collaboration between the device provider and the user of the smart medication device to update the device with the new medicines (CMO 9). This could be done effectively if the device provider regularly reminds the user to report any changes in medication regimens.

Prescription of temporary medicines such as antibiotics can also lead to reduction in medication adherence due to required changes in medication intake habits. An intervention study that provided patients with smart pillboxes before hospital discharge [35] did not include temporary medications inside the pillbox and instead provided individuals with text message reminders to take their temporary medications. Similarly, workshop participants also suggested that the temporary medications should be kept outside the smart medication devices as they could interfere with monitoring of the adherence of other medicines, and a note could be included inside the device to remind them to take their temporary medications. Alternatively, workshop participants suggested that these medicines could be kept inside the device, and the device provider could remotely reset the device after the individual no longer needs to take them (CMO 10).

Workshop participants also suggested that pro re nata (PRN) or "when required" medicines such as laxatives should be kept outside the smart medication devices as they can interfere with monitoring of other medicines. Participants also suggested that hand-written notes reminding individuals to check if they need to take the PRN medicines should be included in the device (CMO 11). Similarly, the intervention study by Shahani et al [35] that provided patients recently discharged from hospital with smart pillboxes also did not include PRN medications inside the pillboxes. Individuals were instead sent text message reminders to check if they needed to take these medications.

### *CMO Derived Only From Workshops*

Workshop participants acknowledged that complex medication regimens could make medication adherence challenging. The smart medication device that they were discussing provided users with reminders to take medications but did not specify the name of the medication and the time to take it. Thus, participants suggested that a written chart of the names of medications and the correct time to take each medicine should be included in the box. This can improve individuals' knowledge of medication regimens and medication adherence. Additionally, the device providers mentioned that they would recommend the individual schedule an appointment with the pharmacist or general practitioner for a potential medication review to simplify medication regimes (CMO 12).

## Theory 3: Enabling Collection of Prescription Medications on Time

### Overview

This theory area includes the CMOs related to ways by which digital interventions can improve collection of prescription medications on time. These CMOs were derived from the literature only. The overarching program theory presented in Figure 4 is based on the CMOs of theory 3 presented in Multimedia Appendix 4.

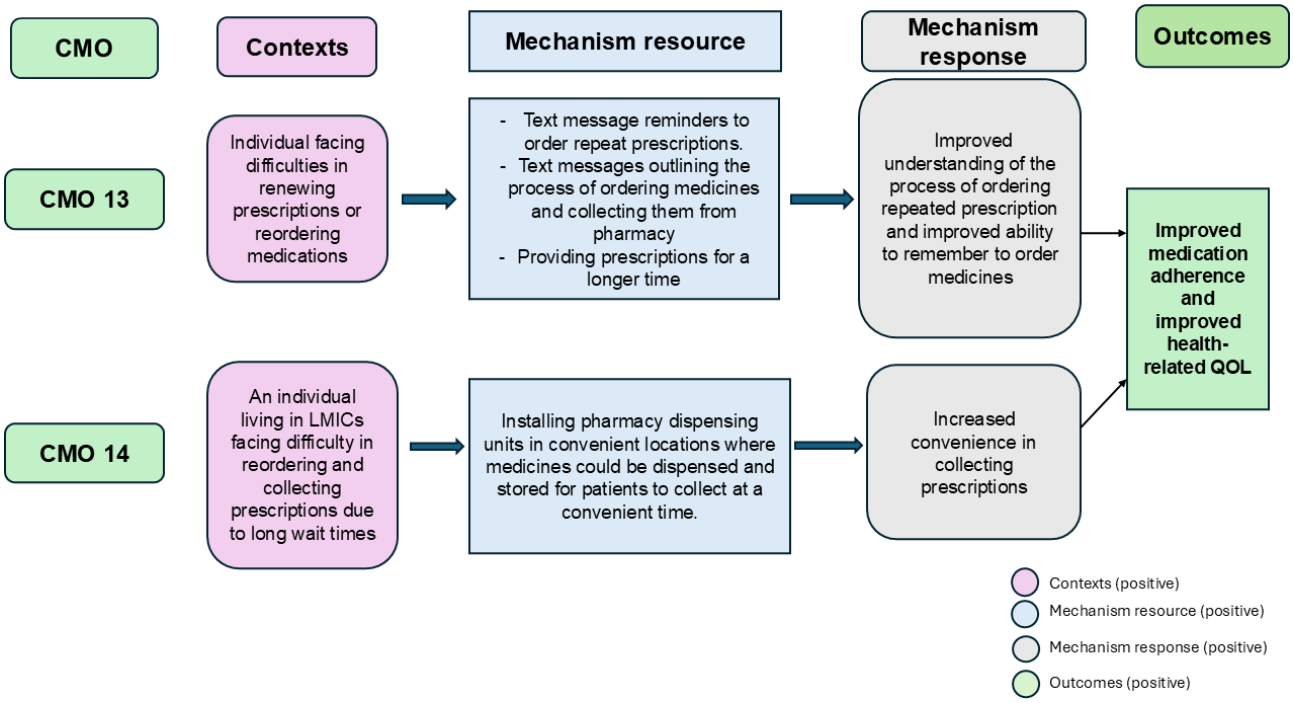
Studies have highlighted the importance of enabling easier access to prescription medications to prevent primary nonadherence. Text message reminders or reminders from mobile apps to order prescription medications on time and outlining the exact process of ordering prescriptions and collecting them can improve medication adherence. This would be helpful for individuals who forget to order prescriptions or who are unable to understand the process of ordering repeated prescriptions. Additionally, in the context in which frequent medication reviews are not required, providing repeat prescriptions for longer periods of time can improve medication adherence as it would reduce the frequency of ordering prescriptions (CMO 13) [48,68,79,80,99].

Mash et al [100] suggested that collecting prescription medications on time could be particularly challenging for individuals living in low- and middle-income countries (LMICs) due to limited human resources and poor infrastructure in primary health care services, resulting in long waiting periods to collect prescriptions. Thus, this study [100] suggested the installation of pharmacy dispensing units which allow remote dispensing of medications (medication dispensing can be requested online by patients) and storage of medicines in password-protected lockers in convenient areas. This can allow patients to collect their medicines at a convenient time and help in improving medication adherence (CMO 14). However, installation of pharmacy dispensing units requires high start-up costs. Additionally, the long-term efficacy, feasibility, and acceptability of these technologies are yet to be determined, as these have been piloted on a very small scale only in South Africa [77,100].



XSL•FO
RenderX

**Figure 4.** The overarching figure showing the CMOs in theory 3: enabling collection of prescription medications on time. CMO: context-mechanism-outcome; LMIC: low- and middle-income country; QOL: quality of life.




## Theory 4: Supporting People With Sensory or Motor Impairments and Unpredictable Health Issues

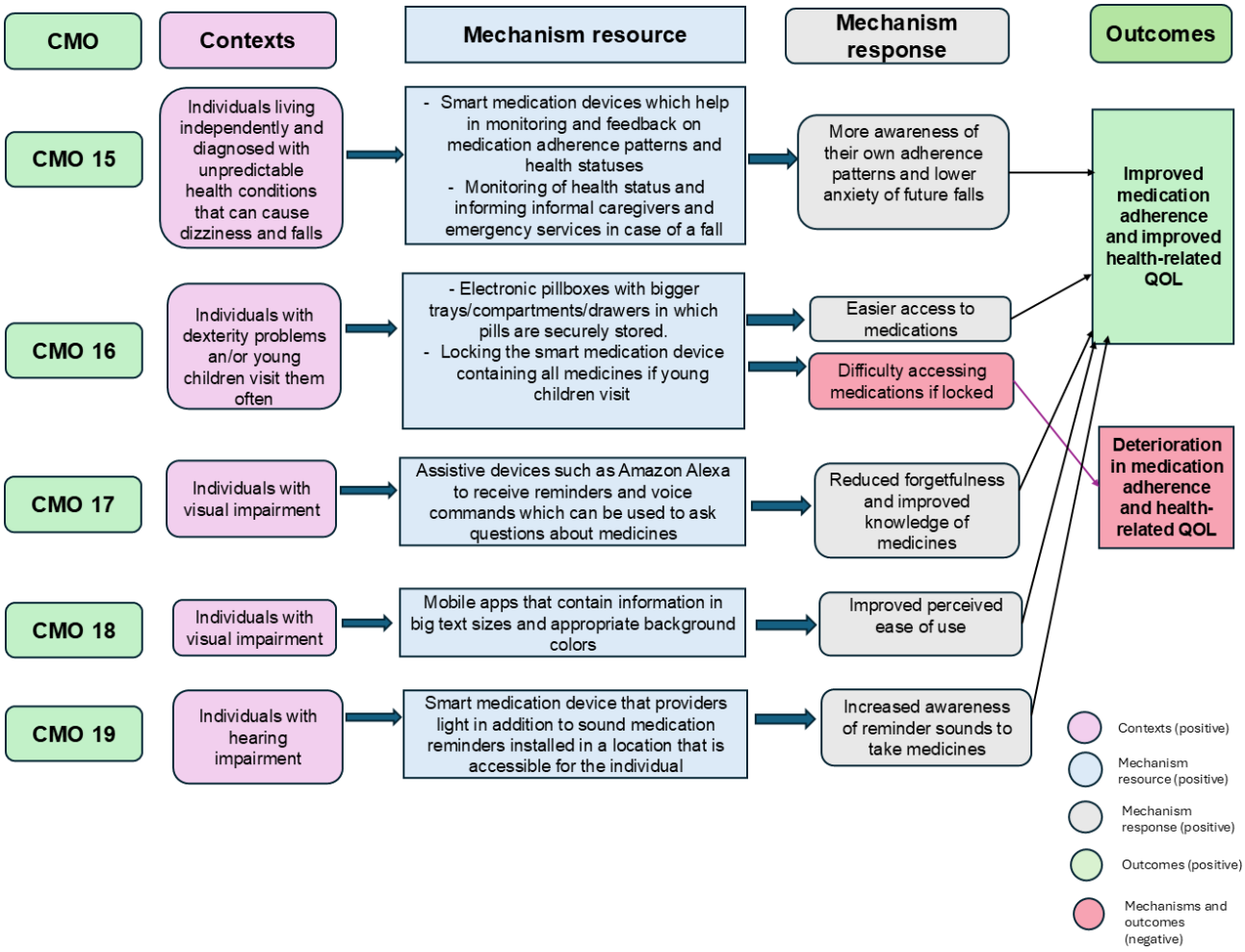
### *Overview*

This theory area includes the CMOs related to ways by which digital medication adherence interventions can improve the adherence of people with unpredictable health conditions and sensory and motor deficits. The overarching program theory presented in Figure 5 is based on the CMOs of theory 4 presented in Multimedia Appendix 4.



XSL•FO
RenderX

**Figure 5.** The overarching figure showing the CMOs in theory 4: supporting people with sensory or motor impairments and unpredictable health issues. CMO: context-mechanism-outcome; QOL: quality of life.




### *CMOs Derived From Both Literature and Workshops*

Smart medication devices that allow health care professionals and device providers to monitor medication adherence and provide feedback on adherence patterns based on the monitoring can improve individuals' adherence by increasing their awareness of their adherence patterns. Changes in the health status of individuals could be detected by observing any changes in the frequency of individuals' use of devices (CMO 15). If an individual misses several doses of medications, a timely intervention could be provided [30,33-35,42,52,56,57,60,61,64,65,73,78,81-83,86,88,95,99,101-103]. Emergency services and informal caregivers could be contacted in the context of severe deterioration of health or a fall. Thus, these devices can also be used for effective telecare monitoring in situations where older adults live alone [52].

However, some individuals with dexterity issues, for example, people with arthritis, may struggle to use devices such as smart pillboxes, as it may be difficult for them to take the medicines out of the containers. Providing these individuals with smart pillboxes with bigger compartments or pill trays could enable easier access to medications [34,42,56,73,74]. Additionally, workshop participants highlighted that if an individual is visited often by young children, the smart medication devices might need to be locked or kept out of reach of children. However, if an individual has poor dexterity and struggles to open locks, they might struggle to unlock the devices to access their medications, resulting in poor medication adherence (CMO 16).

### *CMOs Derived Only From the Literature*

Several studies recommended that the mobile apps aiming to support individuals' medication adherence should include bigger text sizes and appropriate backgrounds to enable individuals with visual impairments to see the content properly (CMO 17) [40,91]. Devices such as Amazon Alexa that provide auditory medication intake reminders can also be helpful in improving medication adherence among individuals with visual impairment (CMO 18). These devices also have voice command features which can be used to ask questions regarding the prescribed medications, which can enable a better understanding of the prescribed medicines, thereby improving medication adherence [41,74].

### *CMO Derived Only From Workshops*

As described earlier (Methods section), the smart medication device discussed in the workshops provided light and sound medication intake reminders. Thus, participants felt that the device was suitable for individuals with visual and hearing impairments. However, participants mentioned that individuals with hearing impairment may not be able to hear the reminder sounds if it is not installed in an accessible location. Thus, installing the device at an accessible location in individuals'



XSL•FO
RenderX

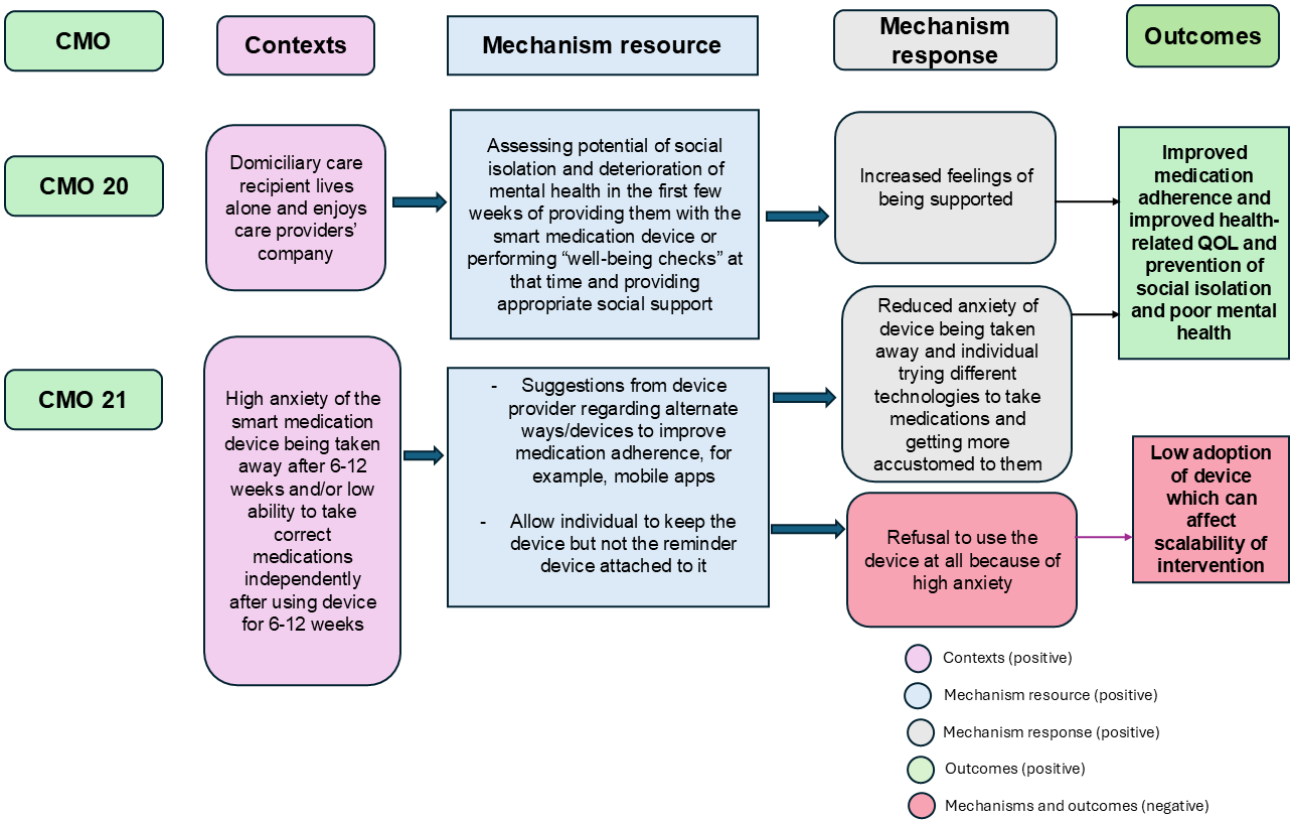
homes was identified as a useful mechanism to improve the adherence of individuals with hearing impairment (CMO 19).

### Theory 5: Supporting Individuals With High Anxiety, Low Medication Self-Efficacy, and Low Social Support

This theory area includes the CMOs related to how the smart medication device that was provided to domiciliary care recipients and hospital discharge patients in the United Kingdom could help in improving the medication adherence of patients with high anxiety, low self-efficacy, and low social support. The CMOs in this section were derived from the workshop discussions only. The overarching program theory presented in Figure 6 is based on the CMOs of theory 5 presented in Multimedia Appendix 4.

Domiciliary care recipients in the United Kingdom who have difficulties with medication adherence are generally visited by care providers to prompt them to take their medications. Therefore, participants discussed that if a domiciliary care recipient lives alone and has low social support, they may enjoy the company of care providers and look forward to their visits. These care recipients may feel isolated if the care providers' visits are reduced when they self-manage their medicines. Thus, participants suggested that the device provider should assess the mental health and the potential for social isolation of these individuals and provide appropriate social support after the device is installed in their homes. This is especially important for individuals with a prior history of mental health diagnoses such as depression (CMO 20).

**Figure 6.** The overarching figure showing the CMOs in theory 5: supporting individuals with high anxiety, low medication self-efficacy, and low social support. CMO: context-mechanism-outcome; QOL: quality of life.




As mentioned in the Methods section, this device was provided to the recipients for only 6 - 12 weeks, and during that time, the device was aimed at helping them to develop effective medication intake habits, after which the device was taken away. Participants mentioned that the removal of the device could cause anxiety among some individuals as they may become dependent on it for maintaining their medication adherence. Participants recommended that users could be allowed to keep the device that stores all the medicines safely, but the electronic reminder, which is used to provide medication intake reminders and monitor individuals' adherence (the costly part of the device), could be taken away by the device provider. This could ensure that all the medicines are stored in the same place and strengthen individuals' habits of taking medications from the same place, resulting in reduced anxiety and maintenance of medication adherence. However, some individuals with high anxiety may refuse to use the device at the outset if they are informed that it would be taken away after 6 - 12 weeks. This could reduce the large-scale adoption and scalability of the intervention. The device provider also mentioned that they could suggest, following removal, alternative devices to maintain their medication adherence, for example, mobile apps. This could help individuals to try different types of technology to manage their medications and improve their comfort with technology (CMO 21).



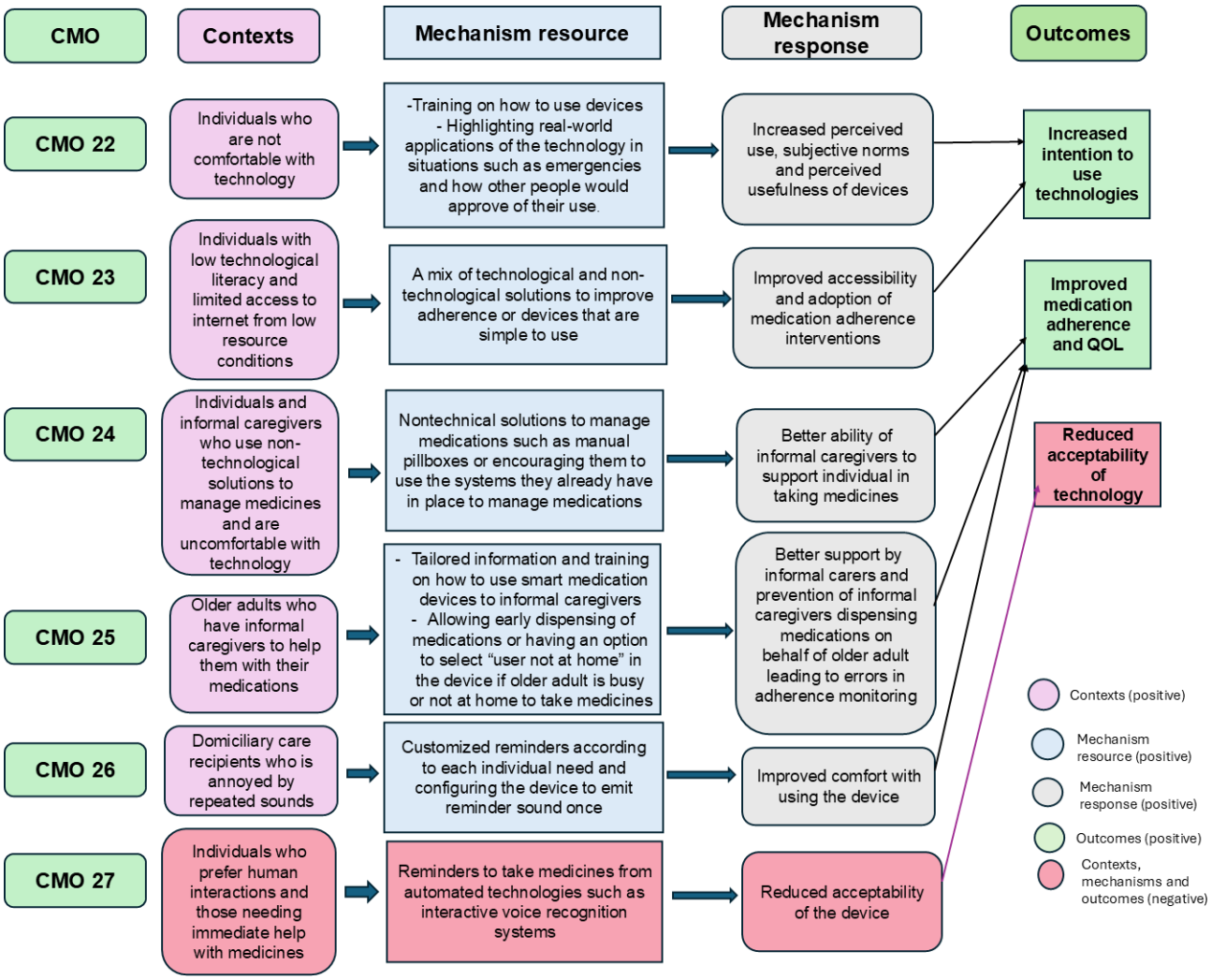
XSL•FO
RenderX

## Theory 6: Supporting Individuals Who Are Uncomfortable With Technology or Certain Aspects of Technology

### *Overview*

This theory area includes CMOs related to mechanisms by which digital medication adherence interventions can improve the adherence of individuals who are uncomfortable with technology or certain aspects of technology, such as repeated reminder sounds for taking medications. The overarching program theory presented in Figure 7 is based on the CMOs of theory 6 presented in Multimedia Appendix 4.

**Figure 7.** The overarching figure showing the CMOs in theory 6: supporting individuals who are uncomfortable with technology or certain aspects of technology. CMO: context-mechanism-outcome; QOL: quality of life.




### *CMOs Derived From Both Literature and Workshops*

Workshop participants mentioned that it was important to provide some training on how to use the smart medication device, especially to individuals who were uncomfortable with technology (CMO 22). This was corroborated by the results of a qualitative study [40] which showed that providing hands-on training on the use of a mobile app intervention in small groups could increase the perceived ease of use of the technology. Additionally, highlighting the usefulness of the mobile app in real-world situations such as health emergencies can increase perceived usefulness. Highlighting how their family members or doctors may approve of their use of the app to improve their adherence could improve subjective norm (individuals' perception of how others would perceive their behavior). According to the Technology Acceptance Model, an increase in perceived usefulness, perceived ease of use, and subjective norms can increase the intention to use the technology to improve medication adherence [40].

Workshop participants also highlighted the importance of providing some training on how to use smart medication devices to informal caregivers or family members so that they can better support the person they care for. Additionally, some of the smart medication dispensers that provide medication intake reminders, allow monitoring of adherence, and dispense medications according to individual doses may need to be adapted for use if informal caregivers support individuals with their medication intake [42]. Some of these devices assume that the medication has been taken by the individual correctly once the device has been opened and medications are dispensed [42]. However, if the individual is busy or not at home, informal caregivers may dispense the medication on behalf of the individual. Therefore, these devices should have an option to select "user not at home" to indicate that the individual is not at home or is busy [42] to



XSL•FO
RenderX

improve the accuracy of medication adherence monitoring (CMO 23).

Some studies suggested that nontechnological solutions or a combination of technological and nontechnological solutions may be more suitable to improve the adherence of individuals living in low-resource conditions and LMICs. This is important as poor infrastructure, low internet access, and technological literacy among individuals are common in these contexts [46,77,104,105]. This was corroborated by workshop discussions as participants noted that one of the advantages of the smart medication device discussed was that it had a built-in SIM card and no Wi-Fi was needed to use it, which made it more accessible for individuals living in low-resource conditions (CMO 24).

### *CMOs Derived Only From the Literature*

A qualitative study conducted with informal caregivers of older adults [47] showed that many informal caregivers already had established nontechnical methods of supporting medication adherence of the person they were caring for (eg, paper medication lists with descriptions of medicines and information about when to take medicines and a box to store all the medicines). Thus, the informal caregivers were unwilling to learn and use technologies such as mobile apps to support medication adherence as they felt that use of these technologies would increase caregiver burden. In these contexts, encouraging individuals to use these nontechnological solutions may be more suitable (CMO 25).

CMO 26 highlighted the low acceptability of automated technologies such as interactive voice recognition systems and automated telephone communication systems to receive medication intake reminders [38,74,85,106] among patients who preferred speaking to a person. These technologies are also unsuitable for patients needing immediate support regarding their health condition or medication intake, and using these technologies can result in lower medication adherence for these individuals [38,106].

### *CMO Derived Only From Workshops*

A few care recipients who participated in the workshops mentioned that repeated medication intake reminder sounds from the device may be annoying, and this could negatively affect medication adherence of individuals if they open the device and take the wrong medications to stop the reminder sounds. Thus, participants suggested that to increase individuals' comfort with using the technology, the reminder sounds should be customized to individual needs, and the frequency of the reminders should be reduced to only once if required (CMO 27).

## Theory 7: Economic, Policy, and Organizational Factors Affecting Implementation of Digital Medication Interventions

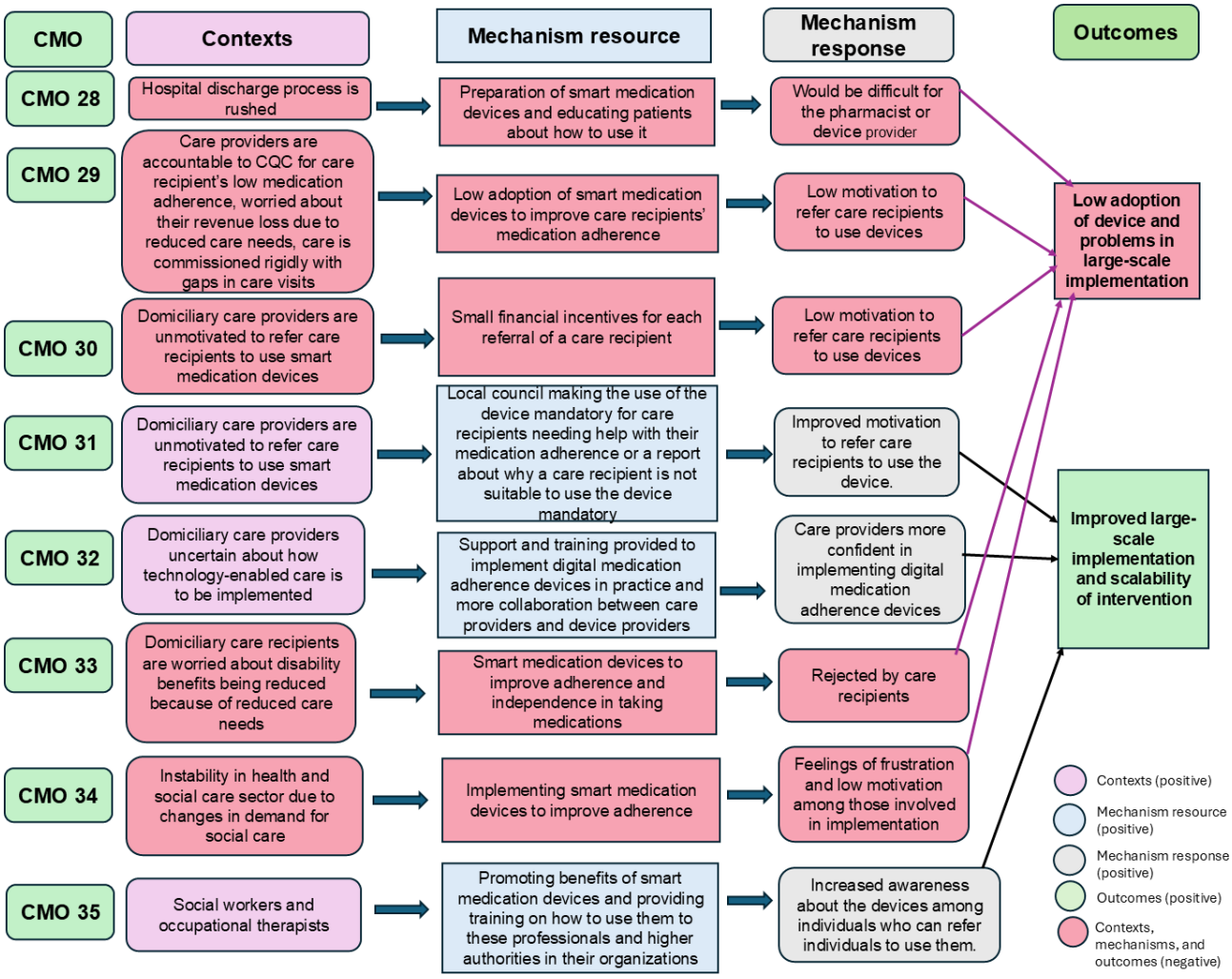
### *Overview*

This theory area includes CMOs related to the wider economic, policy, or workplace organizational factors that affect large-scale implementation of digital medication adherence interventions. The overarching program theory presented in Figure 8 is based on the CMOs of theory 7 presented in Multimedia Appendix 4.



XSL•FO
RenderX

**Figure 8.** The overarching figure showing the CMOs in theory 7: economic, policy, and organizational factors affecting implementation of digital medication interventions. CMO: context-mechanism-outcome; CQC: Care Quality Commission; NHS: National Health Service.




### *CMO Derived Only From the Literature*

An intervention study which provided patients with smart pillboxes before hospital discharge [35] highlighted that there were several barriers at the organizational level relating to implementation of the intervention. The rushed nature of hospital discharge made it difficult for pharmacists to prepare the pillbox and demonstrate the use of the pillbox before patients were discharged. Thus, the authors recommended that organizational changes such as early communication between prescribing clinicians and pharmacists preparing the pillboxes and early provision of the discharge medications are required for successful implementation in the context of hospital discharge (CMO 28) [35].

### *CMO Derived Only From Workshops*

The workshop participants who were involved in implementing the device highlighted several barriers to the large-scale implementation of the device. The number of individuals who were being referred to use the device from domiciliary care provider organizations and short-term re-enablement services that provide care to patients recently discharged from hospitals in the United Kingdom was very low [110].

Participants felt that the scalability of the intervention was potentially reduced because obtaining referrals from domiciliary care organizations was particularly difficult because care providers were anxious about potential revenue loss for them. Their anxiety was further increased because of a recent increase in the number of care providers locally as compared to the demand for their services (CMO 29). Additionally, care providers have to undertake more administrative tasks to identify the care recipients who were suitable to use the device and refer them to the team implementing the device. Care providers would also have to report any lapses in medication adherence among care recipients to the Care Quality Commission that regulates the domiciliary care services in the United Kingdom [111]. Care providers were worried that they may be held accountable for any lapses in medication adherence of care recipients. Care provision is generally commissioned in the United Kingdom with rigid specifications of tasks and times of visits, with several gaps between the care visits. Early morning and late-night visits are common. Care providers prefer to undertake some medication prompting visits during the rest of the day to increase the hours of their work shift and their financial compensation.

The device provider mentioned that they offered care organizations small financial incentives (£25 - £30; a currency exchange rate of £1=US $1.35 was applicable) for each referral.



XSL•FO
RenderX

However, this did not improve care providers' motivation to refer care recipients to use the device because of fear of long-term revenue loss. A higher value of financial incentives would have also limited the cost savings that could have been achieved from this intervention (CMO 30).

Thus, participants felt that the local councils that commission care services should make the use of the device a part of usual care for care recipients needing support with their medication adherence. If an individual is not suitable for using the device, a report from care providers outlining the reasons for that should be required. Participants felt that this may have improved the implementation of the intervention (CMO 31). Additionally, the local councils were aiming to increase technology-enabled care provision [111], but care providers were uncertain about how to implement it in their care provision. Thus, participants felt that local councils should support care providers to implement technologies such as smart medication devices in their care provision and encourage more collaborations between them and care providers (CMO 32).

Participants also highlighted that care recipients may not be motivated to use the device because they could be anxious about the potential reduction of their disability benefits if they require less support for daily living tasks and mobility due to a reduction in the care providers' visits (CMO 33) [112].

Overall, participants felt that the instability in the health and social care sector in the United Kingdom in recent years contributed to the problems with implementation. The number of care providers available and demand for their services fluctuated significantly in the previous years due to the national government's efforts to improve their recruitment and retention. This resulted in a reduced demand for their services in the local area when the project was delivered [111,113]. Thus, participants suggested that future projects should allocate enough time at the beginning to assess contextual factors and to form effective partnerships with organizations that can help improve the implementation of devices (CMO 34). Participants also suggested that the benefits of this device should be promoted to professionals such as social workers and occupational therapists as they regularly assess the care needs of individuals. It was also important to promote the device among the higher authorities of the organizations in which these professionals work to get leadership support for the implementation (CMO 35).

## Discussion

### Summary of Findings

Thirty-five CMOs in 7 theory areas were identified. CMOs within theory 1 demonstrated ways in which devices can help establish and maintain medication intake routines, such as by providing medication intake reminders. Theory 2 presented CMOs related to solving medication-related barriers to adherence, for example, self-monitoring of medication adherence to reduce double or missed doses. Theory 3 presented CMOs related to enabling collection of prescription medications on time, for example, by sending text message reminders to collect prescriptions on time. Theory 4 included CMOs related to improving the medication adherence of individuals with sensory and motor impairments and unpredictable health conditions, for example, encouraging individuals with visual impairments to use assistive devices with voice command features such as Alexa to receive medication intake reminders. Theory 5 presented CMOs related to customizing smart medication devices according to the needs of individuals with high anxiety and low social support and low medication self-efficacy, such as assessing the mental health of individuals who are socially isolated and providing them with appropriate social support. Theory 6 presented CMOs related to supporting individuals who are uncomfortable with technology, for example, providing training on how to use smart medication devices. Theory 7 presented CMOs related to policy and economic factors that were barriers to the large-scale implementation of smart medication devices, such as a lack of suitable incentives for care providers to refer domiciliary care recipients to use the device.

CMOs were derived from both the literature and workshops (14/35, 40%), with a proportion derived only from the literature (7/35, 20%) and a proportion only from the workshops (14/35, 40%). CMOs derived from workshops provided important information about implementing smart medication devices in the United Kingdom and in the context of domiciliary care and hospital discharge. CMO 14 (enabling easier access to prescriptions in LMICs) was not adequately supported by high-quality evidence, and further research is needed to support it. After assessing the "rigor" at a program theory level, the authors excluded a CMO that was initially developed regarding ingestible digital pills for monitoring and providing feedback on medication adherence because these have recently been discontinued due to low acceptability and adoption [114].

### Comparison With Prior Work and Implications for Practice and Policy

The results of the current realist review are consistent with those of a qualitative study [114] in which patients diagnosed with cardiovascular diseases acknowledged that medication intake reminders and reminders for keeping water ready to take with medications helped in maintaining effective medication intake habits. Similarly, the importance of tailoring the timing and frequency of the reminders according to individual preferences to prevent "alert fatigue" and annoyance related to receiving repeated reminders was also highlighted [115]. The importance of action planning related to medication intake by linking it with daily routine tasks such as mealtimes [116,117] was also supported by previous studies.

Previous studies have also found that self-monitoring and monitoring and feedback on adherence by health care providers could help in improving adherence, including reducing the possibility of missed doses [118,119]. Smart pillboxes or medication dispensers that allow monitoring of adherence have also been recommended by previous studies as an effective device for telecare monitoring of older adults and individuals diagnosed with unpredictable health conditions living independently. However, these devices should also be tailored for people with visual, hearing and dexterity problems [120]. A related issue shown in a previous qualitative study highlighted



XSL•FO
RenderX

the importance of locking smart medication devices to prevent any interference by children [121], but it should still be easy to open for people with dexterity problems. However, all digital medication interventions should address privacy concerns [42,45,49,59,88,103,104]. Terms and conditions educating users on data protection and ways to protect their privacy, password protection of data, and an option to consent to the collection and use of their personal data should be included [64,108].

There was only 1 study included in the review that reported factors influencing large-scale implementation and scalability of a digital medication adherence intervention [35]. However, the workshop discussions identified several economic and policy factors that significantly hampered the large-scale implementation of the smart medication device that was provided to the patients at hospital discharge and domiciliary care recipients in the ICB region, United Kingdom. Examples of this included a recent decline in demand for adult domiciliary care services resulting in care providers' fears of revenue loss, care providers' fears of being held responsible for medication errors by care recipients [111], and care recipients' fears of losing disability benefits due to reduction in their care [112]. As a higher value of financial incentives would have also limited the cost savings from this intervention, policy changes to make the use of the device the usual care for individuals requiring medication prompting visits from care providers should be considered right from the beginning of any person's care plan.

### Directions for Future Research

Only 1 study [35] included in this review focused on re-establishing effective medication intake habits after hospital discharge; this was also the only study included in the review that assessed and reported factors affecting the large-scale implementation of the intervention. While the program theories derived from the workshop discussions filled this gap to an extent, future digital medication adherence intervention studies should explore and report the wider social, policy, and economic contextual factors that affect the scalability of the interventions. This is important as insights into factors affecting scalability of previous interventions can help in developing more effective implementation strategies for future interventions [122]. Further research is also needed to explore health system and policy changes needed to effectively implement medication adherence devices in the context of hospital discharge.

Lessons learned from the cocreation workshops were that it was essential for future studies to assess the wider social, economic, and policy factors that can potentially affect the scalability before starting the implementation of an intervention. It was also important to allocate enough time at the beginning of projects to develop effective relationships with key professionals in the community who can support the effective implementation of smart medication devices. Effective implementation strategies could be co-designed with key professionals who could support large-scale implementation of digital medication adherence interventions in the community at the beginning of the project [123].

Additionally, assessment of "richness" of the studies showed that details regarding the content and delivery methods of medication adherence interventions were lacking in some studies, which somewhat hindered the exploration of additional CMOs [9,29,30,35,43,44,49,53,56,58,62,69-72,75-77,80-84,86,88,92-94,96-98,100,103-106,108]. Thus, it is recommended that future medication adherence intervention studies describe the context, content, and delivery of the interventions in detail [123]. Moreover, the wide variation in the methods used in assessing adherence made it difficult for reviews to compare results among studies [83,106]. Some studies assessed adherence using only self-report measures, which are prone to bias and overestimation of adherence. There are drawbacks in using objective methods as well; for example, many digital tools can only assess when the packaging is opened and closed but do not measure if the medication has been ingested. Therefore, assessing adherence by using a combination of objective and self-report measures has been recommended for a comprehensive assessment of adherence [124,125].

### Strengths and Limitations

A key strength of this study was its rigorous adherence to Medical Research Council guidelines [123] to uncover the contexts and mechanisms related to the efficacy or inefficacy of digital medication adherence interventions using realist synthesis. Additionally, cocreation workshops with diverse stakeholders of a project aiming to implement a digital medication adherence intervention led to the identification of CMOs of the intervention, which led to triangulation of the data. It also led to a better understanding of the implementation of such interventions in the community in the United Kingdom. Given the rapidly changing social policy and social care environment, a further strength was the use of the repeated workshops at approximately a 3-month interval with the same participants, which also gained insights as the implementation of the device evolved.

However, a limitation is that the CMOs developed from the workshop discussions could not be empirically verified and further refined using the cost-effectiveness and the effectiveness data of the device in improving medication adherence and quality of life because of the challenges in recruiting participants to use it. Nevertheless, one of the strengths of this study was that it highlighted that, in the future, when designing potential projects for interventions of this nature, it is important to assess the wider contexts and design some kind of flexibility into any project plans so that if wider contextual changes emerge, there is scope for the project to change as well.

This study focused on individuals who had adequate capacity to self-manage medication intake, and so papers which focused on medication adherence of people with severe mental illnesses, dementia, and substance use were excluded. Thus, the program theories derived from this review may not be applicable for medication adherence interventions for these populations. Only program theories associated with improving medication adherence among unintentionally nonadherent individuals were included; therefore, these program theories may not be generalizable for interventions designed for individuals who are intentionally nonadherent.



XSL•FO
RenderX

## Conclusions

Digital interventions are commonly used to improve medication adherence; however, their efficacy may be limited because of a lack of effective tailoring for different populations and contexts [8]. Thus, this realist review addresses this gap in the literature by providing recommendations for tailoring digital medication adherence interventions effectively. Medication intake reminders and reminders for keeping water ready before medication intake, self-monitoring, and monitoring and feedback on medication adherence by device providers or health care professionals were the main active ingredients of interventions to reduce unintentional nonadherence across all theory areas. That said, intervention developers and implementers should adapt these active ingredients for specific contexts and patient subgroups by considering the recommendations provided in the CMO configurations identified in this study. This is important as the same active ingredients will not be effective across contexts because of how contextual factors or patients' responses interact with the mechanism resources. For example, a person with visual impairment may be better supported by a device that provides auditory medication intake reminders (eg, Alexa) rather than pillboxes or mobile apps. Some CMOs in this paper suggest different delivery methods of medication adherence interventions in specific contexts; for example, CMO 11 suggests not including PRN medicines in medication devices and instead providing text message reminders or including written notes in the device as reminders to take them. In these situations, the intervention implementers should choose the delivery methods that align with the patients' needs and preferences [123] and avoid implementing the mechanisms that can potentially result in unintended or negative consequences.

The review also provided insights into the economic, policy, and organizational factors affecting implementation and ways to overcome them. However, further studies exploring these factors are required, especially in the context of domiciliary care provision and hospital discharge. Future interventions should report intervention content and delivery in detail and codevelop effective implementation strategies with professionals who can support the large-scale implementation [123].

## Acknowledgments

We are grateful to the Lancashire and South Cumbria Integrated Care Board for its support throughout this project. We thank all the participants who took part in the logic model workshops. The authors state that there was no use of generative AI technology in the generation of text, figures, or other informational content of this manuscript.

## Funding

This work was funded through the Adult Social Care Technology Fund, provided by the Department of Health and Social Care (DHSC) and distributed by NHS (National Health Service) England, who offer strategic oversight and implementation support to funded projects. The views expressed in this paper are those of the authors and not necessarily those of the funder.

## Data Availability

The datasets used and/or analyzed during this study are available from the corresponding author upon reasonable request.

## Authors' Contributions

Conceptualization: SS (equal), CH (equal)
Data curation – SS
Formal analysis – SS
Funding acquisition – CH (lead), YL (supporting)
Investigation – SS (equal), YL (equal), CH (equal)
Methodology – SS (equal), YL (equal), CH (equal)
Project administration – SS (equal), CH (equal)
Supervision – YL (supporting), CH (lead)
Visualization – SS (lead), YL (supporting),
CH (supporting)Writing – original draft – SS
Writing – review & editing – YL (equal), CH (equal)

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Cocreation workshop details.
[DOCX File, 658 KB - i-jmr_v15i1e89803_app1.docx ]

Multimedia Appendix 2



XSL•FO
RenderX

Realist review search strategies.
[DOCX File, 22 KB - i-jmr_v15i1e89803_app2.docx ]

Multimedia Appendix 3
Data extraction sheet.
[DOCX File, 57 KB - i-jmr_v15i1e89803_app3.docx ]

Multimedia Appendix 4
Context-mechanism-outcomes and quotes supporting the context-mechanism-outcomes.
[DOCX File, 49 KB - i-jmr_v15i1e89803_app4.docx ]

## References


1. Aremu TO, Oluwole OE, Adeyinka KO, Schommer JC. Medication adherence and compliance: recipe for improving patient outcomes. Pharmacy (Basel) 2022 Aug 28;10(5):106. [doi: 10.3390/pharmacy10050106] [Medline: 36136839]
2. Adherence to long-term therapies: evidence for action. : World Health Organization; 2003 URL: http://apps.who.int/iris/bitstream/10665/42682/1/9241545992.pdf [accessed 2026-08-11]
3. Hartch CE, Dietrich MS, Stolldorf DP. Effect of a medication adherence mobile phone app on medically underserved patients with chronic illness: preliminary efficacy study. JMIR Form Res 2023 Dec 11;7:e50579. [doi: 10.2196/50579] [Medline: 38079192]
4. Vrijens B, De Geest S, Hughes DA, et al. A new taxonomy for describing and defining adherence to medications. Br J Clin Pharmacol 2012 May;73(5):691-705. [doi: 10.1111/j.1365-2125.2012.04167.x] [Medline: 22486599]
5. Gellad WF, Grenard J, McGlynn EA. A review of barriers to medication adherence: a framework for driving policy options. : RAND Corporation; 2009 URL: https://www.rand.org/pubs/technical_reports/TR765.html [accessed 2026-08-11]
6. Yap AF, Thirumoorthy T, Kwan YH. Systematic review of the barriers affecting medication adherence in older adults. Geriatr Gerontol Int 2016 Oct;16(10):1093-1101. [doi: 10.1111/ggi.12616] [Medline: 26482548]
7. Wang W, Lau Y, Loo A, Chow A, Thompson DR. Medication adherence and its associated factors among Chinese community-dwelling older adults with hypertension. Heart Lung 2014;43(4):278-283. [doi: 10.1016/j.hrtlng.2014.05.001] [Medline: 24856232]
8. Konstantinou P, Kassianos AP, Georgiou G, et al. Barriers, facilitators, and interventions for medication adherence across chronic conditions with the highest non-adherence rates: a scoping review with recommendations for intervention development. Transl Behav Med 2020 Dec 31;10(6):1390-1398. [doi: 10.1093/tbm/ibaa118] [Medline: 33231691]
9. Armitage LC, Kassavou A, Sutton S. Do mobile device apps designed to support medication adherence demonstrate efficacy? A systematic review of randomised controlled trials, with meta-analysis. BMJ Open 2020 Jan 30;10(1):e032045. [doi: 10.1136/bmjopen-2019-032045] [Medline: 32005778]
10. Anglada-Martinez H, Riu-Viladoms G, Martin-Conde M, Rovira-Illamola M, Sotoca-Momblona JM, Codina-Jane C. Does mHealth increase adherence to medication? Results of a systematic review. Int J Clin Pract 2015 Jan;69(1):9-32. [doi: 10.1111/ijcp.12582] [Medline: 25472682]
11. Linn AJ, Vervloet M, van Dijk L, Smit EG, Van Weert JCM. Effects of eHealth interventions on medication adherence: a systematic review of the literature. J Med Internet Res 2011 Dec 5;13(4):e103. [doi: 10.2196/jmir.1738] [Medline: 22138112]
12. Waldron C, Hughes J, Wallace E, Cahir C, Bennett K. Contexts and mechanisms relevant to general practitioner (GP) based interventions to reduce adverse drug events (ADE) in community dwelling older adults: a rapid realist review. HRB Open Res 2022;5:53. [doi: 10.12688/hrbopenres.13580.2] [Medline: 38283368]
13. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review--a new method of systematic review designed for complex policy interventions. J Health Serv Res Policy 2005 Jul;10 Suppl 1:21-34. [doi: 10.1258/1355819054308530] [Medline: 16053581]
14. Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. RAMESES publication standards: realist syntheses. J Adv Nurs 2013 May;69(5):1005-1022. [doi: 10.1111/jan.12095] [Medline: 23356726]
15. Pawson R, Tilley N. An introduction to scientific realist evaluation. In: Chelimsky E, Shadish WR, editors. Evaluation for the 21st Century: A Handbook: SAGE Publications; 1997:405-418. [doi: 10.4135/9781483348896.n29]
16. Dalkin SM, Greenhalgh J, Jones D, Cunningham B, Lhussier M. What's in a mechanism? Development of a key concept in realist evaluation. Implement Sci 2015 Apr 16;10:49. [doi: 10.1186/s13012-015-0237-x] [Medline: 25885787]
17. Hunter R, Gorely T, Beattie M, Harris K. Realist review. Int Rev Sport Exercise Psychol 2022 Dec 31;15(1):242-265. [doi: 10.1080/1750984X.2021.1969674]
18. Griffiths S, Weston L, Morgan-Trimmer S, et al. Engaging stakeholders in realist programme theory building: insights from the prospective phase of a primary care dementia support study. Int J Qual Methods 2022 Apr;21:16094069221077521. [doi: 10.1177/16094069221077521]




XSL•FO
RenderX

19. Saha DS, Latham DY, Holland PC. Co-creating logic models of the context-mechanism-outcome configurations for a digital intervention to support medication adherence. JMIR Formative Research. Preprint posted online on Jun 3, 2026. [doi: 10.2196/preprints.103497]
20. Marshall P, Booth M, Coole M, et al. Understanding the impacts of online mental health peer support forums: realist synthesis. JMIR Ment Health 2024 May 9;11:e55750. [doi: 10.2196/55750] [Medline: 38722680]
21. The automated pill dispenser project. end project evaluation report. : Improvement and Efficiency West Midlands & NHS Midlands and East; 2012 Mar URL: https://www.pivotell.co.uk/downloads/WMPDReport.pdf?srsltid=AfmBOorU5zwnJAURuZPEX9y596KEow7l-lUycx-JFmWNsVRU0aAi81BD [accessed 2026-08-11]
22. Mason M, Cho Y, Rayo J, Gong Y, Harris M, Jiang Y. Technologies for medication adherence monitoring and technology assessment criteria: narrative review. JMIR mHealth uHealth 2022 Mar 10;10(3):e35157. [doi: 10.2196/35157] [Medline: 35266873]
23. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210. [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
24. Wilson NM, March LM, Sambrook PN, Hilmer SN. Medication safety in residential aged-care facilities: a perspective. Ther Adv Drug Saf 2010 Oct;1(1):11-20. [doi: 10.1177/2042098610381418] [Medline: 25083192]
25. Magola-Makina E, Abuzour AS, Ashcroft DM, Dunlop J, Brown P, Keers RN. Exploring the challenges to safer prescribing and medication monitoring in prisons: a qualitative study with health care staff. PLoS One 2022;17(11):e0275907. [doi: 10.1371/journal.pone.0275907] [Medline: 36327312]
26. Haynes RB, Wang E, Da Mota Gomes M. A critical review of interventions to improve compliance with prescribed medications. Patient Educ Couns 1987 Oct;10(2):155-166. [doi: 10.1016/0738-3991(87)90095-4]
27. Schnorrerova P, Matalova P, Wawruch M. Medication adherence and intervention strategies: why should we care. Bratisl Med J 2025;126(7):1196-1206. [doi: 10.1007/s44411-025-00227-0]
28. Tyndall J. AACODS Checklist.: Flinders University; 2010 Nov. URL: https://fac.flinders.edu.au/dspace/api/core/bitstreams/e94a96eb-0334-4300-8880-c836d4d9a676/content [accessed 2026-08-19]
29. Sarzynski E, Decker B, Thul A, et al. Beta testing a novel smartphone application to improve medication adherence. Telemed J E Health 2017 Apr;23(4):339-348. [doi: 10.1089/tmj.2016.0100] [Medline: 27564971]
30. Brondi R, Bannò F, Bendinelli S, et al. Drugs don't work in patients who don't take them: Dr. Drin, the new ICT paradigm for chronic therapies. Stud Health Technol Inf 2013;191:100-104. [Medline: 23792852]
31. Arain MA, Ahmad A, Chiu V, Kembel L. Medication adherence support of an in-home electronic medication dispensing system for individuals living with chronic conditions: a pilot randomized controlled trial. BMC Geriatr 2021 Jan 14;21(1):56. [doi: 10.1186/s12877-020-01979-w] [Medline: 33446126]
32. Mira JJ, Navarro I, Botella F, et al. A Spanish pillbox app for elderly patients taking multiple medications: randomized controlled trial. J Med Internet Res 2014 Apr 4;16(4):e99. [doi: 10.2196/jmir.3269] [Medline: 24705022]
33. He J, Chen J, Li Q, et al. Application of family-involved smart medication management system in rural-dwelling middle-aged and older adult participants with chronic diseases: management of chronic diseases in rural areas. Medicine (Baltimore) 2022;101(45):e31662. [doi: 10.1097/MD.0000000000031662]
34. Faisal S, Ivo J, Tennant R, et al. Integration of a smart multidose blister package for medication intake: a mixed method ethnographic informed study of older adults with chronic diseases. PLoS One 2022;17(1):e0262012. [doi: 10.1371/journal.pone.0262012] [Medline: 35061773]
35. Shahani A, Nieva HR, Czado K, et al. An electronic pillbox intervention designed to improve medication safety during care transitions: challenges and lessons learned regarding implementation and evaluation. BMC Health Serv Res 2022 Oct 30;22(1):1304. [doi: 10.1186/s12913-022-08702-y] [Medline: 36309744]
36. Fallah M, Yasini M. A medication reminder mobile app: does it work for different age ranges. Stud Health Technol Inf 2017;235:68-72. [Medline: 28423757]
37. Wiecek E, Torres-Robles A, Cutler RL, Benrimoj SI, Garcia-Cardenas V. Impact of a multicomponent digital therapeutic mobile app on medication adherence in patients with chronic conditions: retrospective analysis. J Med Internet Res 2020 Aug 12;22(8):e17834. [doi: 10.2196/17834] [Medline: 32784183]
38. Reidel K, Tamblyn R, Patel V, Huang A. Pilot study of an interactive voice response system to improve medication refill compliance. BMC Med Inf Decis Making 2008 Oct 9;8:46. [doi: 10.1186/1472-6947-8-46] [Medline: 18845004]
39. Castillo AF, Davis AL, Fischhoff B, Krishnamurti T. Digital medicines for adherence support: a conceptual framework and qualitative study of adherence among chronically ill patients. Health Inf J 2021;27(5):14604582211059463. [doi: 10.1177/14604582211059463] [Medline: 34825829]
40. Park DY, Goering EM, Head KJ, Bartlett Ellis RJ. Implications for training on smartphone medication reminder app use by adults with chronic conditions: pilot study applying the technology acceptance model. JMIR Form Res 2017 Nov 10;1(1):e5. [doi: 10.2196/formative.8027] [Medline: 30684397]
41. Balasubramanian GV, Beaney P, Chambers R. Digital personal assistants are smart ways for assistive technology to aid the health and wellbeing of patients and carers. BMC Geriatr 2021 Nov 15;21(1):643. [doi: 10.1186/s12877-021-02436-y] [Medline: 34781881]




XSL•FO
RenderX

42. Ahmad A, Chiu V, Arain MA. Users' perceptions of an in-home electronic medication dispensing system: a qualitative study. Med Devices (Auckl) 2020;13:31-39. [doi: 10.2147/MDER.S241062] [Medline: 32104106]
43. Anglada-Martínez H, Martin-Conde M, Rovira-Illamola M, et al. Feasibility and preliminary outcomes of a web and smartphone-based medication self-management platform for chronically Ill patients. J Med Syst 2016 Apr;40(4):99. [doi: 10.1007/s10916-016-0456-y] [Medline: 26872781]
44. Stecher C, Linnemayr S, Reaven P, Cloonan S, Huckfeldt P. Cost savings from an mHealth tool for improving medication adherence. Am J Managed Care 2024 Oct 1;30(10):e289-e296. [doi: 10.37765/ajmc.2024.89621] [Medline: 39467173]
45. Russell AM, Smith SG, Bailey SC, et al. Older adult preferences of mobile application functionality supporting medication self-management. J Health Commun 2018;23(12):1064-1071. [doi: 10.1080/10810730.2018.1554728] [Medline: 30526400]
46. Hincapie AL, Gupta V, Brown SA, Metzger AH. Exploring perceived barriers to medication adherence and the use of mobile technology in underserved patients with chronic conditions. J Pharm Pract 2019 Apr;32(2):147-153. [doi: 10.1177/0897190017744953] [Medline: 29212410]
47. LaValley SA, Brady LA, Nguyen M, Jacobs D. Family caregivers of older adults: strategies, tools, and technologies for medication management. J Gerontol Soc Work 2023;66(2):291-302. [doi: 10.1080/01634372.2022.2100856] [Medline: 35850634]
48. Donovan G, Hall N, Smith F, Ling J, Wilkes S. Two-way automated text messaging support from community pharmacies for medication taking in multiple long-term conditions: human-centered design with nominal group technique development study. JMIR Form Res 2022 Dec 21;6(12):e41735. [doi: 10.2196/41735] [Medline: 36542458]
49. Backes C, Moyano C, Rimaud C, Bienvenu C, Schneider MP. Digital medication adherence support: could healthcare providers recommend mobile health apps? Front Med Technol 2020;2:616242. [doi: 10.3389/fmedt.2020.616242] [Medline: 35047896]
50. Peng Y, Wang H, Fang Q, et al. Effectiveness of mobile applications on medication adherence in adults with chronic diseases: a systematic review and meta-analysis. J Managed Care Spec Pharm 2020 Apr;26(4):550-561. [doi: 10.18553/jmcp.2020.26.4.550] [Medline: 32223596]
51. Pouls BPH, Vriezekolk JE, Bekker CL, et al. Effect of interactive eHealth interventions on improving medication adherence in adults with long-term medication: systematic review. J Med Internet Res 2021 Jan 8;23(1):e18901. [doi: 10.2196/18901] [Medline: 33416501]
52. Gackowski M, Jasińska-Stroschein M, Osmałek T, Waszyk-Nowaczyk M. Innovative approaches to enhance and measure medication adherence in chronic disease management: a review. Med Sci Monit 2024 Jul 16;30:e944605. [doi: 10.12659/MSM.944605] [Medline: 39012851]
53. Zhou TT, Wang R, Gu SJ, et al. Effectiveness of mobile medical apps in ensuring medication safety among patients with chronic diseases: systematic review and meta-analysis. JMIR mHealth uHealth 2022 Nov 22;10(11):e39819. [doi: 10.2196/39819] [Medline: 36413386]
54. Wiecek E, Tonin FS, Torres-Robles A, Benrimoj SI, Fernandez-Llimos F, Garcia-Cardenas V. Temporal effectiveness of interventions to improve medication adherence: a network meta-analysis. PLoS One 2019;14(3):e0213432. [doi: 10.1371/journal.pone.0213432] [Medline: 30861014]
55. Andrikopoulou E, Scott P, Herrera H, Good A. What are the important design features of personal health records to improve medication adherence for patients with long-term conditions? A systematic literature review. BMJ Open 2019 Sep 26;9(9):e028628. [doi: 10.1136/bmjopen-2018-028628] [Medline: 31558449]
56. Miguel-Cruz A, Bohórquez AF, Parra PAA. What does the literature say about using electronic pillboxes for older adults? A systematic literature review. Disabil Rehabil Assistive Technol 2019 Nov;14(8):776-787. [doi: 10.1080/17483107.2018.1508514] [Medline: 30451543]
57. Emadi F, Ghanbarzadegan A, Ghahramani S, Bastani P, Baysari MT. Factors affecting medication adherence among older adults using tele-pharmacy services: a scoping review. Arch Public Health 2022 Aug 31;80(1):199. [doi: 10.1186/s13690-022-00960-w] [Medline: 36042508]
58. Colon C, Salas P, Díaz M, et al. Patient assistance programs and technology in medication adherence. Innovations Pharm 2020;11(2):10. [doi: 10.24926/iip.v11i2.3269] [Medline: 34007605]
59. Eggerth A, Hayn D, Schreier G. Medication management needs information and communications technology‐based approaches, including telehealth and artificial intelligence. Brit J Clin Pharmacol 2020 Oct;86(10):2000-2007. [doi: 10.1111/bcp.14045]
60. Hersberger KE, Boeni F, Arnet I. Dose-dispensing service as an intervention to improve adherence to polymedication. Expert Rev Clin Pharmacol 2013 Jul;6(4):413-421. [doi: 10.1586/17512433.2013.811829] [Medline: 23927669]
61. Billingsley L, Carruth A. Use of technology to promote effective medication adherence. J Contin Educ Nurs 2015 Aug;46(8):340-342. [doi: 10.3928/00220124-20150721-12] [Medline: 26247653]
62. Thakkar J, Kurup R, Laba TL, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
63. Hugtenburg JG, Timmers L, Elders PJ, Vervloet M, van Dijk L. Definitions, variants, and causes of nonadherence with medication: a challenge for tailored interventions. Patient Preference Adherence 2013;7:675-682. [doi: 10.2147/PPA.S29549] [Medline: 23874088]




XSL•FO
RenderX

64. Liu X, Varshney U. Mobile health: a carrot and stick intervention to improve medication adherence. Decis Support Syst 2020 Jan;128:113165. [doi: 10.1016/j.dss.2019.113165]
65. Marengoni A, Monaco A, Costa E, et al. Strategies to improve medication adherence in older persons: consensus statement from the Senior Italia Federanziani Advisory Board. Drugs Aging 2016 Sep;33(9):629-637. [doi: 10.1007/s40266-016-0387-9] [Medline: 27444182]
66. Bingham JM, Black M, Anderson EJ, et al. Impact of telehealth interventions on medication adherence for patients with type 2 diabetes, hypertension, and/or dyslipidemia: a systematic review. Ann Pharmacother 2021 May;55(5):637-649. [doi: 10.1177/1060028020950726] [Medline: 32815400]
67. Kardas P, Bago M, Barnestein-Fonseca P, et al. Reimbursed medication adherence enhancing interventions in 12 European countries: current state of the art and future challenges. Front Pharmacol 2022;13:944829. [doi: 10.3389/fphar.2022.944829] [Medline: 36034792]
68. Torres-Robles A, Wiecek E, Tonin FS, Benrimoj SI, Fernandez-Llimos F, Garcia-Cardenas V. Comparison of interventions to improve long-term medication adherence across different clinical conditions: a systematic review with network meta-analysis. Front Pharmacol 2018;9:1454. [doi: 10.3389/fphar.2018.01454] [Medline: 30618748]
69. Tantray J, Patel A, Wani SN, Kosey S, Prajapati BG. Prescription precision: a comprehensive review of intelligent prescription systems. Curr Pharm Des 2024;30(34):2671-2684. [doi: 10.2174/0113816128321623240719104337] [Medline: 39092640]
70. Sun S, Simonsson O, McGarvey S, Torous J, Goldberg SB. Mobile phone interventions to improve health outcomes among patients with chronic diseases: an umbrella review and evidence synthesis from 34 meta-analyses. Lancet Digital Health 2024 Nov;6(11):e857-e870. [doi: 10.1016/S2589-7500(24)00119-5] [Medline: 39332937]
71. Pirri S, Lorenzoni V, Turchetti G. Scoping review and bibliometric analysis of Big Data applications for medication adherence: an explorative methodological study to enhance consistency in literature. BMC Health Serv Res 2020 Jul 24;20(1):688. [doi: 10.1186/s12913-020-05544-4] [Medline: 32709237]
72. Schwartz JK, Grogan KA, Mutch MJ, et al. Intervention to improve medication management: qualitative outcomes from a phase I randomized controlled trial. Am J Occup Ther 2017;71(6):7106240010p1-7106240010p10. [doi: 10.5014/ajot.2017.021691] [Medline: 29135431]
73. Haugh KH. Medication adherence in older adults: the pillbox half full. Nurs Clin North Am 2014 Jun;49(2):183-199. [doi: 10.1016/j.cnur.2014.02.006] [Medline: 24846467]
74. Zullig LL, Ryan JS, Hayden BB. Applying technology to medication management and adherence. In: Marsch L, Lord S, Dallery J, editors. Behavioral Health Care and Technology: Using Science-Based Innovations to Transform Practice: Oxford University Press; 2014:81-94. [doi: 10.1093/med/9780199314027.003.0007]
75. Austin RR, Hull S. The power of mobile health technologies and prescribing apps. Comput Inf Nurs 2014 Nov;32(11):513-515. [doi: 10.1097/CIN.0000000000000120] [Medline: 25391038]
76. Viswanathan M, Golin CE, Jones CD, et al. Closing the quality gap: revisiting the state of the science (vol. 4: medication adherence interventions: comparative effectiveness). Evidence Rep Technol Assess (Full Rep) 2012 Sep(208.4):1-685. [Medline: 24422970]
77. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health 2014 Feb 21;14:188. [doi: 10.1186/1471-2458-14-188] [Medline: 24555733]
78. Patient adherence to medicines. Academy of Medical Sciences. 2015. URL: https://acmedsci.ac.uk/policy/policy-projects/patient-adherence-to-medicines [accessed 2025-09-05]
79. Khan R, Socha-Dietrich K. Investing in medication adherence improves health outcomes and health system efficiency. : Organisation for Economic Co-operation and Development; 2018 URL: https://www.oecd.org/en/publications/investing-in-medication-adherence-improves-health-outcomes-and-health-system-efficiency_8178962c-en.html [accessed 2026-08-11]
80. HTA policy and methods review: emerging health technologies. : Australian Government Department of Health and Aged Care; 2024 Jul 10 URL: https://www.health.gov.au/resources/publications/hta-policy-and-methods-review-emerging-health-technologies?language=en [accessed 2026-08-11]
81. Medication safety in polypharmacy: technical report. : World Health Organization; 2019 URL: https://www.who.int/publications/i/item/medication-safety-in-polypharmacy-technical-report [accessed 2026-08-11]
82. Cross AJ, Elliott RA, Petrie K, Kuruvilla L, George J. Interventions for improving medication-taking ability and adherence in older adults prescribed multiple medications. Cochrane Database Syst Rev 2020 May 8;5(5):CD012419. [doi: 10.1002/14651858.CD012419.pub2] [Medline: 32383493]
83. Nieuwlaat R, Wilczynski N, Navarro T, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014 Nov 20;2014(11):CD000011. [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
84. Coulter A, Ellins J. Patient-focused interventions: a review of the evidence.: The Health Foundation; 2006 Aug. URL: https://www.health.org.uk/reports-and-analysis/reports/patient-focused-interventions [accessed 2026-08-19]
85. Doctor's orders: why adherence is critical to improving health outcomes across the life course. : International Longevity Centre UK; 2022 Apr 7 URL: https://ilcuk.org.uk/wp-content/uploads/2022/04/ILC-Doctors-Orders.pdf [accessed 2026-08-19]




XSL•FO
RenderX

86. YOURmeds for medication support in long-term conditions. : National Institute for Health and Care Excellence; 2023 Jun 9 URL: https://www.nice.org.uk/advice/mib289/resources/yourmeds-for-medication-support-in-longterm-conditions-pdf-2285965883209669 [accessed 2026-08-19]
87. Badawy SM, Shah R, Beg U, Heneghan MB. Habit strength, medication adherence, and habit-based mobile health interventions across chronic medical conditions: systematic review. J Med Internet Res 2020 Apr 28;22(4):e17883. [doi: 10.2196/17883] [Medline: 32343250]
88. Schneider M, Gertsch A, Bugnon O. Cyberhealth serving to support individual intake of medication. Swiss Med Wkly 2013 Jul 28;143(3132):w13827. [doi: 10.4414/smw.2013.13827] [Medline: 23925717]
89. Goradia S, Holland R, Alexander S, Greenbaum D, Chen T, Aslani P. A new age intervention to support medication adherence. Res Soc Administrative Pharm 2021 Jun;17(6):1204-1207. [doi: 10.1016/j.sapharm.2020.07.038] [Medline: 32811756]
90. Park LG, Howie-Esquivel J, Dracup K. A quantitative systematic review of the efficacy of mobile phone interventions to improve medication adherence. J Adv Nurs 2014 Sep;70(9):1932-1953. [doi: 10.1111/jan.12400] [Medline: 24689978]
91. Wang T, Huang YM, Chan HY. Exploration of features of mobile applications for medication adherence in Asia: narrative review. J Med Internet Res 2024 Nov 8;26:e60787. [doi: 10.2196/60787] [Medline: 39514859]
92. Jiménez-Chala EA, Durantez-Fernández C, Martín-Conty JL, Mohedano-Moriano A, Martín-Rodríguez F, Polonio-López B. Use of mobile applications to increase therapeutic adherence in adults: a systematic review. J Med Syst 2022 Oct 22;46(12):87. [doi: 10.1007/s10916-022-01876-2] [Medline: 36272023]
93. Bakema R, Smirnova D, Biri D, Kocks JWH, Postma MJ, de Jong LA. The use of eHealth for pharmacotherapy management with patients with respiratory disease, cardiovascular disease, or diabetes: scoping review. J Med Internet Res 2023 Sep 26;25:e42474. [doi: 10.2196/42474] [Medline: 37751232]
94. Lammila-Escalera E, Greenfield G, Pan Z, Nicholls D, Majeed A, Hayhoe B. Interventions to improve medication adherence in adults with mental-physical multimorbidity in primary care: a systematic review. Br J Gen Pract 2024 Jul;74(744):e442-e448. [doi: 10.3399/BJGP.2023.0406] [Medline: 38429109]
95. Doggrell SA. Adherence to medicines in the older-aged with chronic conditions: does intervention by an allied health professional help? Drugs Aging 2010 Mar 1;27(3):239-254. [doi: 10.2165/11532870-000000000-00000] [Medline: 20210369]
96. Choi JY, Choi H, Seomun G, Kim EJ. Mobile-application-based interventions for patients with hypertension and ischemic heart disease: a systematic review. J Nurs Res 2020 Oct;28(5):e117. [doi: 10.1097/JNR.0000000000000382] [Medline: 32501962]
97. Artificial intelligence in health care: benefits and challenges of technologies to augment patient care. : US Government Accountability Office; 2020 Nov 30 URL: https://www.gao.gov/products/gao-21-7sp [accessed 2026-08-11]
98. Lancaster K, Abuzour A, Khaira M, et al. The use and effects of electronic health tools for patient self-monitoring and reporting of outcomes following medication use: systematic review. J Med Internet Res 2018 Dec 18;20(12):e294. [doi: 10.2196/jmir.9284] [Medline: 30563822]
99. Aungst TD. Reevaluating medication adherence in the era of digital health. Expert Rev Med Devices 2021 Dec;18(sup1):25-35. [doi: 10.1080/17434440.2021.2019012] [Medline: 34913793]
100. Mash R, Christian C, Chigwanda RV. Alternative mechanisms for delivery of medication in South Africa: a scoping review. S Afr Fam Pract 2001 Aug 24;63(1):e1-e8. [doi: 10.4102/safp.v63i1.5274]
101. Espírito-Santo M, Santos S, Estêvão MD. Digital solutions available to be used by informal caregivers, contributing to medication adherence: a scoping review. Pharmacy (Basel) 2024 Jan 23;12(1):20. [doi: 10.3390/pharmacy12010020] [Medline: 38392927]
102. Schneider MP, Krummenacher I, Figueiredo H, Marquis J, Bugnon O. Adherence: a review of education, research, practice and policy in Switzerland. Pharm Pract (Granada) 2009 Apr;7(2):63-73. [doi: 10.4321/s1886-36552009000200001] [Medline: 25152780]
103. Perry C, Chhatralia K, Damesick D, Hobden S, Volpe L. Behavioural insights in health care: nudging to reduce inefficiency and waste.: The Health Foundation; 2015 Dec. URL: https://www.health.org.uk/reports-and-analysis/reports/behavioural-insights-in-health-care [accessed 2026-08-11]
104. Yasmin F, Banu B, Zakir SM, Sauerborn R, Ali L, Souares A. Positive influence of short message service and voice call interventions on adherence and health outcomes in case of chronic disease care: a systematic review. BMC Med Inf Decis Making 2016 Apr 22;16:46. [doi: 10.1186/s12911-016-0286-3] [Medline: 27106263]
105. Tolley A, Hassan R, Grewal K, et al. Interventions to promote medication adherence for chronic diseases in India: a systematic review. Clin Med 2023 Nov;23(6):72-73. [doi: 10.7861/clinmed.23-6-s72]
106. Posadzki P, Mastellos N, Ryan R, et al. Automated telephone communication systems for preventive healthcare and management of long-term conditions. Cochrane Database Syst Rev 2016 Dec 14;12(12):CD009921. [doi: 10.1002/14651858.CD009921.pub2] [Medline: 27960229]
107. Donovan G, Hall N, Ling J, Smith F, Wilkes S. Influencing medication taking behaviors using automated two-way digital communication: a narrative synthesis systematic review informed by the Behavior Change Wheel. Br J Health Psychol 2022 Sep;27(3):861-890. [doi: 10.1111/bjhp.12580] [Medline: 35080811]




XSL•FO
RenderX

108. Dada S, Dalkin S, Gilmore B, Hunter R, Mukumbang FC. Applying and reporting relevance, richness and rigour in realist evidence appraisals: advancing key concepts in realist reviews. Res Synth Methods 2023 May;14(3):504-514. [doi: 10.1002/jrsm.1630] [Medline: 36872619]
109. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
110. Intermediate care allocation team. NHS East Lancashire Hospitals. URL: https://elht.nhs.uk/services/intermediate-care-allocation-team [accessed 2026-08-20]
111. Lancashire's market position statement for adult social care. : Lancashire County Council; 2019 URL: https://www.caremanagementmatters.co.uk/wp-content/uploads/2019/09/The-Lancashire-Market-Position-Statement-issues-for-providers-and-commissioners.pdf [accessed 2026-08-19]
112. PIP handbook. GOV.UK. 2024 Nov 25. URL: https://www.gov.uk/government/publications/personal-independence-payment-fact-sheets/pip-handbook#changes-in-circumstances-needs-and-other-notifiable-changes [accessed 2026-08-11]
113. Skills for care report: international recruitment helps social care workforce grow – but domestic recruitment and retention struggles persist. Skills for Care. 2024 Oct 10. URL: https://www.skillsforcare.org.uk/news-and-events/news/skills-for-care-report-international-recruitment-helps-social-care-workforce-grow-but-domestic-recruitment-and-retention-struggles-persist [accessed 2026-08-19]
114. Kardas P. From non-adherence to adherence: can innovative solutions resolve a longstanding problem? Eur J Intern Med 2024 Jan;119:6-12. [doi: 10.1016/j.ejim.2023.10.012] [Medline: 37848351]
115. Park LG, Ng F, Handley MA. The use of the Capability-Opportunity- Motivation Behavior (COM-B) model to identify barriers to medication adherence and the application of mobile health technology in adults with coronary heart disease: a qualitative study. PEC Innov 2023 Dec 15;3:100209. [doi: 10.1016/j.pecinn.2023.100209] [Medline: 37753273]
116. Hoo ZH, Gardner B, Arden MA, et al. Role of habit in treatment adherence among adults with cystic fibrosis. Thorax 2019 Feb;74(2):197-199. [doi: 10.1136/thoraxjnl-2017-211453] [Medline: 29886416]
117. Alhomoud FK, Alwohaibi LW, Aljarrash K, et al. Evaluating strategies for enhancing medication adherence in the Kingdom of Saudi Arabia (KSA): a cross-sectional study. Patient Preference Adherence 2024;18:2469-2480. [doi: 10.2147/PPA.S499795] [Medline: 39669314]
118. Religioni U, Barrios-Rodríguez R, Requena P, Borowska M, Ostrowski J. Enhancing therapy adherence: impact on clinical outcomes, healthcare costs, and patient quality of life. Medicina (Kaunas) 2025 Jan 17;61(1):153. [doi: 10.3390/medicina61010153] [Medline: 39859135]
119. van Heuckelum M, van den Ende CHM, Houterman AEJ, Heemskerk CPM, van Dulmen S, van den Bemt BJF. The effect of electronic monitoring feedback on medication adherence and clinical outcomes: a systematic review. PLoS One 2017;12(10):e0185453. [doi: 10.1371/journal.pone.0185453] [Medline: 28991903]
120. Gargioni L, Fogli D, Baroni P. A systematic review on pill and medication dispensers from a human-centered perspective. J Healthcare Inf Res 2024 Jun;8(2):244-285. [doi: 10.1007/s41666-024-00161-w] [Medline: 38681758]
121. Shannon EM, Mueller SK, Schnipper JL. Patient, caregiver, and clinician experience with a technologically enabled pillbox: a qualitative study. ACI open 2023 Jul;07(2):e61-e70. [doi: 10.1055/s-0043-1775970]
122. Zamboni K, Schellenberg J, Hanson C, Betran AP, Dumont A. Assessing scalability of an intervention: why, how and who? Health Policy Plann 2019 Sep 1;34(7):544-552. [doi: 10.1093/heapol/czz068] [Medline: 31365066]
123. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. Int J Nurs Stud 2024 Jun;154:104705. [doi: 10.1016/j.ijnurstu.2024.104705] [Medline: 38564982]
124. Basu S, Garg S, Sharma N, Singh MM. Improving the assessment of medication adherence: challenges and considerations with a focus on low-resource settings. Tzu Chi Med J 2019;31(2):73-80. [doi: 10.4103/tcmj.tcmj_177_18] [Medline: 31007485]
125. Lam WY, Fresco P. Medication adherence measures: an overview. Biomed Res Int 2015;2015:217047. [doi: 10.1155/2015/217047] [Medline: 26539470]


## Abbreviations

**CMO:** context-mechanism-outcome
**ICB:** integrated care board
**LMIC:** low- and middle-income country
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**PRN:** pro re nata
**PROSPERO:** The International Prospective Register of Systematic Reviews
**RAMESES:** Realist and Meta-Narrative Evidence Synthesis: Evolving Standards



XSL•FO
RenderX

*Edited by M Balcarras; submitted 18.Dec.2025; peer-reviewed by A Borek, P Cook, YH Kwan; revised version received 24.Jul.2026; accepted 24.Jul.2026; published 28.Aug.2026.*

*Please cite as:*
*Saha S, Latham Y, Holland C*
*Contexts and Mechanisms Related to the Efficacy of Digital Medication Adherence Interventions to Support Medication Adherence Among Adults With Chronic Diseases: Realist Review*
*Interact J Med Res 2026;15:e89803*
*URL: https://www.i-jmr.org/2026/1/e89803*
*doi:10.2196/89803*

© Srinjaya Saha, Yvonne Latham, Carol Holland. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 28.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Enhancing Healthy Longevity: Scoping Review of Practices and Interventions

Angela Yee Man Leung[1,2], PhD; Karen Siu Lan Cheung[1,2], PhD; Kai-ling Ou[1,2], PhD; Ivy Yan Zhao[1,2], PhD; Min Qiu[2], MA; Ka Ming Wan[3], MPH; Jed Montayre[1,2,4], PhD; Yaqi Huang[1,2], PhD

[1]WHO Collaborating Centre for Community Health Services, School of Nursing, The Hong Kong Polytechnic University, 5th Floor, Core G, Hung Hom, Kowloon, Hong Kong, China (Hong Kong)

[2]School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[3]University of British Columbia, Vancouver, BC, Canada

[4]JBI Hong Kong Centre of Evidence-based Healthcare Excellence, School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

**Corresponding Author:**
Angela Yee Man Leung, PhD
WHO Collaborating Centre for Community Health Services, School of Nursing, The Hong Kong Polytechnic University, 5th Floor, Core G, Hung Hom, Kowloon, Hong Kong, China (Hong Kong)

## Abstract

**Background:** Over the past 2 decades, concerns have arisen about the distinction between health span and lifespan, highlighting that longevity does not necessarily equate to good health, a concept often referred to as “healthy longevity.” While various strategies have been explored to promote healthy aging and achieve healthy longevity, it remains uncertain which practices are most effective.

**Objective:** This scoping review identifies the existing and emerging practices and interventions that promote healthy longevity, identifies the key components of these practices and interventions, and considers how stakeholders contribute to these practices and interventions.

**Methods:** A scoping review of the literature was conducted using Arksey and O’Malley’s 6-stage framework. The Joanna Briggs Institute Population-Concept-Context framework was used to define the eligibility criteria and select studies reporting practices or interventions aimed at achieving a long health span or promoting healthy longevity carried out in the community. Data were manually extracted by 2 independent reviewers to detail the characteristics of these practices and interventions and guide the data charting process and narrative synthesis. Six databases (PubMed, Web of Science, Embase, Scopus, CINAHL, and Google Scholar) were searched for academic papers published between January 2010 and February 2025.

**Results:** A total of 21 studies met the inclusion criteria. Most studies (n=15, 71%) were published after 2020 and were predominantly conducted in high-income settings (n=17, 81%) across North America, Europe, and Asia. Four study types were identified: interventional (n=7, 33%), intervention development (n=4, 19%), association (n=5, 24%), and descriptive (n=5, 24%). Interventional and intervention development studies primarily described multicomponent programs targeting individual and social determinants of health and generally reported beneficial effects on physical, cognitive, and psychosocial outcomes. Association studies linked micro-, meso-, and macro-level factors (eg, nutrition, household expenditure, housing quality, health insurance, and public financing policy) to healthy longevity indicators. Descriptive studies highlighted themes across the micro and meso levels, including finance, physical activity, mental and spiritual health, digital literacy, independent living, safety, social support, and health care support.

**Conclusions:** The identified community-based strategies, practices, and policies that extend health span represent the joint efforts of multiple stakeholders and disciplines. The implementation of these practices and policies is worthy of being supported. More studies in diverse socioeconomic contexts are needed.

***(Interact J Med Res 2026;15:e94085)*** doi:10.2196/94085

**KEYWORDS**

healthy longevity; strategies; practices; policies; community-based; review



XSL•FO
RenderX

## Introduction

### Background

The concept of health span has increasingly attracted considerable attention within communities as longevity rises; however, not all individuals experience good health in their later years. Health span is defined as the duration of life spent in good health, encompassing psychological well-being, functional independence, cognitive integrity, and the absence of overt disease and disability [1]. These conditions are generally perceived positively within communities, as older adults tend to use fewer resources, such as health care services, rehabilitation, and social care support. An extended health span is highly esteemed due to its association with overall well-being, enhanced quality of life, and increased levels of life satisfaction in late adulthood [1,2].

In the past 2 decades, concerns have emerged regarding the distinction between health span and lifespan, emphasizing that longevity does not inherently equate to health, often termed "healthy longevity." Some researchers argue that the gap between lifespan and health span is widening [3]. Lifespan refers to the duration of an individual's life from birth to death, focusing solely on the number of years lived, irrespective of health conditions. Although lifespan has increased, many individuals spend the latter decade of their lives contending with significant health challenges [3]. Consequently, these individuals experience what is termed a "long lifespan but short health span," suffering the consequences of prolonged ill health that necessitates intensive care or rehabilitation. This disparity between health span and lifespan is increasingly recognized as a global phenomenon. Recent data indicate that the health span-lifespan gap has expanded among 183 World Health Organization (WHO) member states, averaging 9.6 years, posing significant challenges to healthy longevity worldwide [3]. Notably, women exhibit a larger health span-lifespan gap compared to men, primarily owing to a greater burden of noncommunicable diseases [3].

To narrow the health span-lifespan gap, governments have sought to promote healthy aging by mobilizing community members and resources to support individuals to lead healthy and independent lives. This initiative aims to address declines in intrinsic capacity and maximize functional ability through interactions with the environment. Key strategies for promoting healthy aging include encouraging regular exercise, maintaining a healthy diet, and increasing social interaction among older adults. The United Nations Decade of Healthy Aging (2021 - 2030) seeks to shift the global focus from mere longevity to healthy longevity [4,5]. Investments in initiatives that support healthy aging are crucial for preventing health deterioration alongside increasing longevity, which can lead to functional dependency and disability. Ultimately, such efforts benefit society by reducing overall health care costs.

Joint efforts from the community to achieve healthy longevity (or a long health span) have been consistently advocated. Some community-based practices are culturally specific and initiated by Indigenous Peoples, rooted in their beliefs and local resources. While these practices may be descriptive and lack thorough efficacy analysis, they should nonetheless be valued and respected. Additionally, initiatives from nonhealth sectors, such as housing and finance, contribute important perspectives to the concept of healthy longevity. However, the most effective practices or interventions to support healthy longevity for community-dwelling older adults—enabling them to maintain health, safety, and independence while engaging in cherished activities—remain unclear. To date, no scoping review has been conducted to examine existing and emerging practices and interventions that promote healthy longevity. This study addresses this critical knowledge gap by identifying useful practices and interventions, thereby providing a theoretical foundation to guide healthy longevity initiatives.

### Objective

This scoping review aims to identify existing and emerging practices and interventions that promote healthy longevity, highlight the key components of these interventions, and describe the actions stakeholders should take for the development and implementation of these practices or interventions in the community.

## Methods

### Design

A scoping review was chosen instead of a systematic review because it is particularly suited to exploring emerging topics, such as practices that promote healthy longevity, for which conceptual and empirical clarity is still limited [6]. This approach provides a comprehensive overview of the existing literature on the subject, reviews evidence rapidly, and identifies knowledge gaps. The scoping review process followed the guidance of the 6-stage framework described by Arksey and O'Malley [7] and further refined by subsequent authors, including (1) identifying the research questions; (2) identifying relevant studies; (3) selecting studies; (4) charting the data; (5) collating, summarizing, and reporting the results; and (6) the consultation exercise. The review protocol was not preregistered. Ethical approval was not required for this review of the existing literature. The review was reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist (Checklist 1). A retrospective protocol was registered on OSF [8].

### Scoping Review Questions

The specific review questions were as follows: (1) What existing or emerging community-based practices and interventions support healthy longevity? (2) What are the key components of these practices and interventions? (3) How do different stakeholders contribute to these practices and interventions? Framework Stage 1: Identifying the Research Questions.

### Inclusion and Exclusion Criteria

Eligibility criteria were defined according to the Joanna Briggs Institute Population-Concept-Context framework [9].

Population: The included studies focused on middle-aged and older populations (aged 50 years and older), irrespective of gender, socioeconomic status, or health conditions.



XSL•FO
RenderX

Concept: The scope of this review was informed by the conceptual distinction between lifespan and healthy longevity and by the WHO's healthy aging framework [10]. Lifespan refers to the duration of an individual's life from birth to death [11], whereas healthy longevity, health span, or healthy lifespan emphasizes the period of good health and performance [12]. WHO defines healthy aging as the process of developing and maintaining the functional ability that enables well-being in older age. This framework emphasizes home, community, and broader societal environments as key contexts that support older adults' functional ability [10].

In line with these frameworks, this review focused on practices, programs, strategies, or interventions intended to enhance health span, healthy aging, healthy longevity, or healthy life expectancy in community or home settings. Studies focusing solely on biological lifespan extension, or antiaging pharmacological or medication treatments were excluded because they did not address the community-based and health span–oriented focus of this review. Similarly, studies targeting a single clinical risk or disease outcome without reference to broader healthy aging or health span–related aims were outside the scope of this review. Such studies were eligible only when they were explicitly framed as contributing to healthy aging, health span, healthy longevity, or healthy life expectancy or when relevant outcomes related to these concepts were reported.

In this review, practices were defined as activities adopted by older individuals or communities, including those shaped by cultural beliefs, values, or everyday routines. Interventions were defined as structured activities, programs, strategies, or initiatives developed based on empirical evidence, theories, or pilot tests in selected populations. Interventions that were under development or being evaluated in trials were also eligible.

Context: All geographic locations and countries across various income levels (low-, middle-, and high-income) were included. Consistent with the WHO healthy aging framework, this review focused solely on strategies, programs, and initiatives that took place in community or home settings. Practices in institutions, including hospitals, clinics, long-term care institutions, or residential care homes, were excluded because resources in these in settings differ from those everyday living environments.

All study designs (such as randomized controlled trials, quasi-experimental studies, cohort studies, observational studies, and qualitative research) were included to capture the breadth of the literature. Only studies published in English between January 2010 and February 2025 were considered for inclusion. These studies had to explicitly indicate that the aims or outcomes of the practices or interventions were related to improving health span, healthy lifespan, healthy longevity, or healthy aging, be published in peer-reviewed English-language journals, and be conducted in community settings. Animal studies and in vitro research were excluded. Framework Stage 2: Identifying Relevant Studies.

## Search Methods

A comprehensive search and review of the literature published between January 2010 and February 2025 were conducted using databases and platforms including PubMed, Web of Science, Embase, Scopus, CINAHL, and Google Scholar, and searched the reference lists of the included studies. Proximity operators were used in the search strategy (eg, longev* was used to search for longevity).

The search strategy was intentionally bounded by the WHO healthy aging framework and the distinction between lifespan and health span. The aim was to identify community- and home-based practices, programs, strategies, and interventions that support health span, healthy aging, or healthy longevity, rather than biological lifespan extension alone. The Boolean search combined 3 concept blocks: (1) healthy longevity–related terms referring to the review of the health span definition [12], (2) intervention or practice-related terms, and (3) settings. The core search string was: (healthspan OR healthsp* OR "healthy aging" OR "healthy aging" OR "healthy ag*" OR "healthy longevity OR "healthy longev*" OR "healthy life expectancy" OR "healthy life exp*") AND (intervention* OR practice* OR program* OR program*) AND (community OR home) NOT (hospital OR clinic OR "institutional setting" OR "residential care homes" OR institutions), including MeSH and keywords. The term "longevity" was retained as a supplementary broad aging term to capture studies in which health span–related outcomes were described. Search strings were adapted to the syntax and field tags of each database. The search was further refined using relevant subject categories, such as Health Care Science Services, Health Policy Services, and Good Health and Wellbeing, wherever it was available in the database. Framework Stage 2: Identifying Relevant Studies.

## Search Selection and Outcome

Following the search, all identified articles were collated in EndNote [13] and uploaded into the online systematic review software Covidence [14]. Two authors independently screened titles and abstracts against the inclusion criteria, followed by full-text review. Any disagreements between the authors at each stage of the selection process were resolved through discussion. If consensus could not be reached, a third author reviewed the study to make a final decision. Studies that did not meet the Population-Concept-Context criteria were excluded. Framework Stage 3: Selecting Studies.

## Data Extraction and Abstraction

After selection, the included articles were extracted by 2 authors using a predesigned data extraction form in Microsoft Excel. Study characteristics included the following details: authors, year, country, income level, study design, age, other conditions, and sample size. Additional details were extracted based on the study type. For interventional studies, we extracted the following information: authors, year, study design, funding of the program, and strategies, practices, or policies related to healthy aging. For intervention development studies, we recorded authors, year, study design, framework, and strategies, practices, or policies. For association studies, we extracted authors, year, study design, outcomes, and significant or nonsignificant factors related to healthy longevity. Additionally, the key themes identified in descriptive studies were categorized into micro-, meso-, and macro-level framework. Any discrepancy between the 2 reviewers was resolved through discussion, or by an additional member of the project team. Formal quality appraisal



XSL•FO
RenderX

of the included studies was not undertaken, consistent with common practice in scoping reviews. These data are presented in the *Results* section. Framework Stage 4: Charting the Data.

### Synthesis

The recurring themes identified in the reviewed studies focused on the key components of the practices (interventions), the individuals involved in their delivery, the strategies employed, and the outcomes observed following implementation. Next, we conducted a narrative synthesis of the charted data. Strategies, practices, policies, and associated factors identified were mapped onto micro-, meso-, and macro-level determinants of healthy longevity to summarize patterns across study types (Framework Stage 5: Collating, Summarizing, and Reporting the Results). Using the micro-, meso-, macro-level framework, we applied established definitions for each level: the micro level is about individuals and one-on-one interactions, the meso level is about the community level, and the macro level is the societal level, including policy [15]. The optional Framework Stage 6 (Consultation Exercise) was not undertaken in this review.

## Results

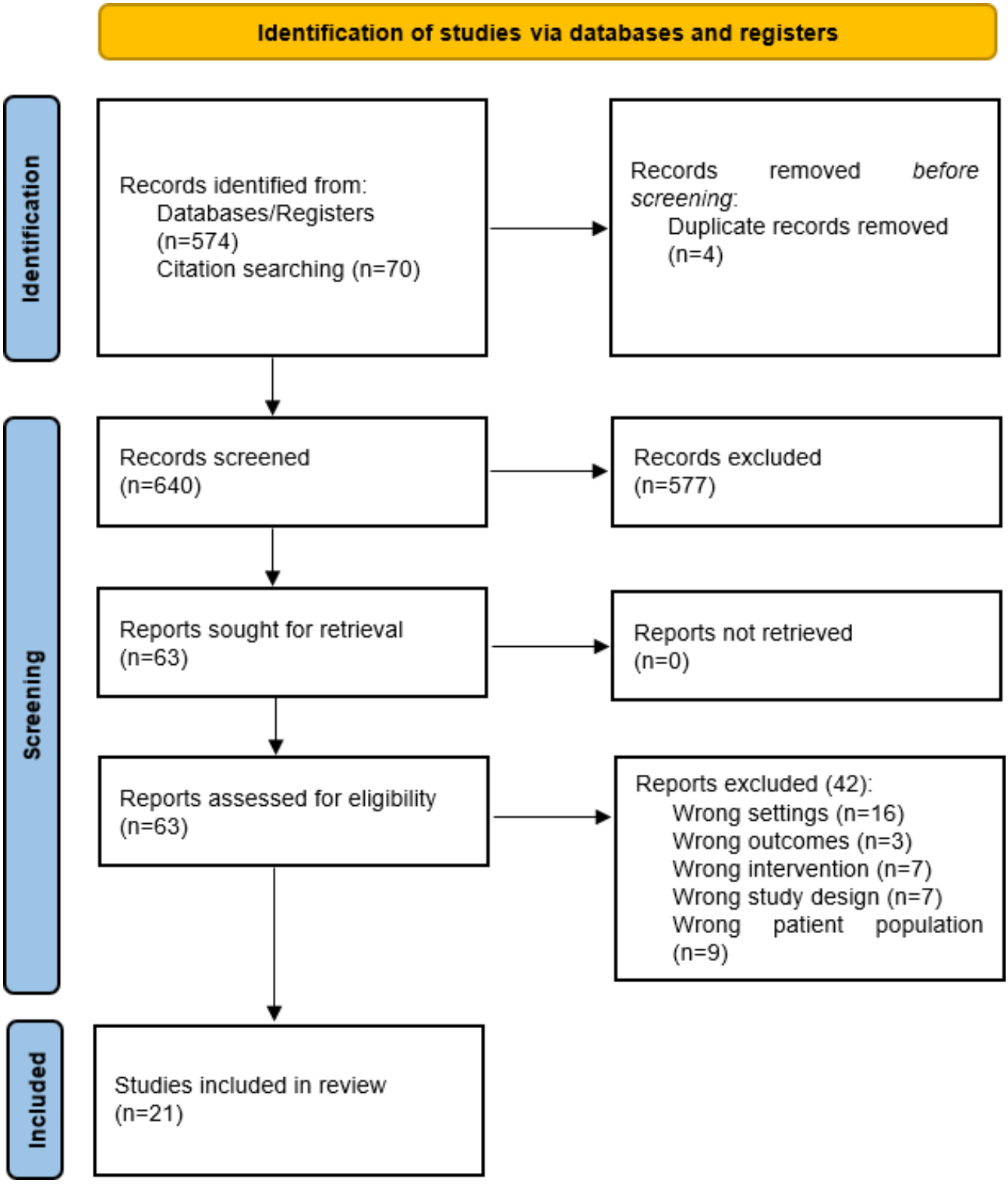
### Study Selection

A total of 644 titles and abstracts were identified and screened to remove duplicates. After title screening, 504 articles were excluded. Sixty-three full-text articles were retrieved, of which 42 were excluded as they did not fulfill the eligibility criteria. Finally, 21 articles were included in this scoping review. Figure 1 details the results of the search, and the inclusion process is reported in full according to the PRISMA-ScR flow diagram [16].



XSL•FO
RenderX

**Figure 1.** PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flowchart of the literature search strategy.




## Study Characteristics

The studies included in this review were published between January 2010 and February 2025, with most published after 2020 (15/21, 71.4%). Most of the studies were conducted in high-income countries or economies (17/21, 81.0%), and the rest were in middle-income countries (4/21, 19.0%), with about one-third in North America (8/21, 38.1%), one-third in Europe (7/21, 33.3%), and Asia (6/21, 28.6%). Four of 21 (19.0%) studies were conducted in rural or low-income communities. Studies conducted in low-income communities mainly focused on basic preventive health risk practices using existing resources, while programs in other affluent settings implemented multidomain approaches. Among the 20 studies recruiting participants, 15 (75%) recruited participants or informants aged 60 years or older, and 10 (50%) studies had additional criteria for target participants: had chronic diseases, lived independently at their own home or lived alone, or lived in low-income areas.



XSL•FO
RenderX

In the review, we identified 4 approaches to the study of interventions for healthy aging: interventional studies (n=7), intervention development studies (n=4), association studies (n=5), and descriptive studies (n=5). First, interventional studies are those that used experimental designs to assess the impact of a specific intervention to improve healthy aging outcomes, including active aging, physical performance, cognitive performance, preventive health behaviors, survival rate, and health service utilization. Second, intervention development studies include those that reported the development of a specific intervention for healthy aging without empirical examination of its impact. Third, association studies used longitudinal studies examining factors associated with measures of healthy aging, such as mortality, disability-adjusted life years (DALYs), Activities of Daily Living/Instrumental Activities of Daily Living, and depression. Fourth, descriptive studies using qualitative methods illuminated participant perspectives and identified themes associated with interventions of healthy aging, such as safety environment, finances, and support systems. Each of these four approaches offers important insights into interventions affecting healthy aging. The sample sizes ranged from 30 to 462 for the 7 interventional studies, 18 to 112 for 4 intervention development studies, and were at least 3000 in 4 of the 5 association studies which had provided the information. For the descriptive studies, the sample sizes ranged from 19 to 47. Details of study characteristics are provided in Table 1.



XSL•FO
RenderX

**Table .** Study characteristics (n=21).

| Authors, year | Country | Income | Study design | Age (y) | Other conditions | Sample size |
|---|---|---|---|---|---|---|
| Interventional studies | | | | | | |
| Davodi et al [17], 2023 | Iran | Middle | RCT[a] | ≥60 | —[b] | 60 |
| Newman et al [18], 2010 | United States | High | Pre-and-post | ≥65 | Low-income high-risk community | 389 |
| Zgibor et al [19], 2017 | United States | High | Cluster RCT | ≥50 | Arthritis and multimorbidity | 462 |
| Lee et al [20], 2022 | Taiwan | High | Secondary data analysis of RCT | ≥65 | At least 3 chronic diseases | 297 |
| Moriyama et al [21], 2023 | Japan | High | Quasi-experimental | ≥65 | Not receiving support/certification for long-term care | 307 |
| Carter et al [22], 2021 | United States | High | Pre-and-post | ≥55 | Live in low-income housing | 30 |
| Lucifora and Villar [23], 2024 | Italy | High | Quasi-experimental | ≥65 | Retirees of a private company | 356 |
| Intervention development studies | | | | | | |
| Joymangul et al [24], 2024 | Italy, Romania, and Portugal | High | Mixed methods observational | No info | — | 112 |
| Orellano-Colón et al [25], 2014 | Puerto Rico | High | Mixed methods | >70 | Older adults who lived alone | Older adults: 8, community key members: 4; researchers: 6 |
| Gwyther et al [26], 2018 | Greece, Italy, the Netherlands, Poland, Portugal, Spain, the United Kingdom | High | Mixed methods | N/A[c] | N/A | 21 project partners |
| Ashikali et al [27], 2024 | Switzerland | High | A participatory approach | ≥65 | — | 36 (older adults: 4; HCPs[d]: 13; SCPs[e]: 9; community representatives: 10) |
| Association studies | | | | | | |
| Nie et al [28], 2021 | China | Middle | Longitudinal | ≥60 | — | 8839 |
| Wu and Grundy [29], 2025 | United Kingdom | High | Longitudinal | ≥65 | — | 4710 |
| Zai [30], 2024 | United States | High | Longitudinal | ≥65 | At risk of requiring long-term care | 3000 |
| Dragos et al [31], 2022 | 26 European countries | High | Longitudinal panel | No info | — | No information |
| McIntyre et al [32], 2021 | United States | High | Longitudinal | ≥18 | — | Training: 2634; validation: 2505 |
| Descriptive studies | | | | | | |
| Ienca et al [33], 2021 | Switzerland | High | Semistructured interviews | ≥65 | Lived independently in their own home | 19 |
| Solhi et al [34], 2022 | Iran | Middle | Semistructured interview | 45 - 59 | — | 21 |
| Puplampu et al [35], 2022 | Canada | High | Semistructured interviews | ≥65 | — | 19 |



XSL•FO
RenderX

| Authors, year | Country | Income | Study design | Age (y) | Other conditions | Sample size |
|---|---|---|---|---|---|---|
| Somrongthong et al [36], 2014 | Thailand | Middle | Semistructured interviews with site observation | ≥65 | Rural community | 47 (older adults: 16; family: 17; volunteers: 11; staff: 3 |
| Bacsu et al [37], 2012 | Canada | High | Ethnographic approach with interviews | ≥65 | Rural community | 42 |

[a]RCT: randomized controlled trial.

[b]Not available.

[c]N/A: not applicable.

[d]HCP: health care professional.

[e]SCP: social care professional.

## Interventional and Intervention Development Studies

Seven studies described specific interventions to improve healthy aging, with various study designs, including RCTs or cluster RCTs (n=3), quasi-experiment with control group (n=2), and a predesign and postdesign (n=2) (Table 2). No study aimed at the macro level, 1 study targeted integration of care at meso level [22], 2 studies targeted lifestyle modifications at the micro level [20,21], and the remaining 4 studies examined multidomain interventions, with most components at the micro level, including physical exercise, nutrition, stress/depression management, and preventive health behaviors, and also one component (social contact/engagement) at the meso level [17-19,23]. Outcome measures were divided into two categories: (1) epidemiological outcomes such as active aging scores [17], survival rate [21], and disability-free life expectancy [23], and (2) physio-psychosocial outcomes: physiological measures [18], cognitive performance [20], social engagement [17,19,23], and mental health [23].



XSL•FO
RenderX

**Table .** Details and outcomes of the interventional studies (n=7)[a].

| Authors, year | Funding of the program | Strategies/practices/policies of healthy aging | Outcomes |
|---|---|---|---|
| Davodi et al [17], 2023 | Research study | Micro<br>• Intervention:<br>  • Nutrition<br>  • Physical activity<br>  • Responsibility for preventive behaviors<br>  • Stress management<br>  • Spiritual aspects<br>• Control:<br>  • Routine care<br>Meso<br>• Communications | • (+) Total active-aging score increased (68.5±3.0 → 85.0±8.3, *P*<.001)<br>• (+) Significant improvements were observed in active-mind maintenance (9.0±2.0 → 13.0±4.0, *P*<.001), physical-functional activity (15.0±4.0 → 16.0±4.0, *P*=.03), social contact (11.5± 4.0 → 16.0±3.3, *P*<.001), productive engagement (13.5±6.0 → 19.0±6.0, *P*<.001), and social-institutional participation (6.0±1.0 → 9.0±4.0, *P*<.001).<br>• (=) No significant difference in agent attitude (13.7±1.60 →13.6±1.60, *P*=.96) |
| Newman et al [18], 2010 | No funding | Micro<br>The 10 keys:<br>• Blood pressure control<br>• Smoking cessation<br>• Immunization<br>• Cancer screening<br>• Regulating blood glucose<br>• Regulating cholesterol<br>• Physical activity<br>• Maintaining healthy bones, joints, and muscles<br>• Combating depression<br>Meso<br>• Promoting social contact | • (↑) Improvements were seen for the proportion of participants meeting goals for low-density lipoprotein cholesterol (+43%), blood pressure control in hypertensives (+17%), blood glucose control in diabetics (+50%), and colon cancer screening (+13%). Among those without prior vaccination, influenza vaccine increased by 25% and pneumonia vaccine by 20% |
| Zgibor et al [19], 2017 | Research grant | Micro<br>• AEEP[b] sessions<br>  • Combining exercise<br>  • Health education on chronic disease prevention and healthy aging<br>• The 10 keys<br>  • Blood pressure control<br>  • Smoking cessation<br>  • Immunization<br>  • Cancer screening<br>  • Regulating blood glucose<br>  • Regulating cholesterol<br>  • Physical activity<br>  • Maintaining healthy bones, joints, and muscles<br>  • Combating depression<br>Meso<br>• Promoting social contact | • (=) Both groups improved, with no significant between-group differences in Short Physical Performance Battery (9.7±2.4 → 9.9 ± 2.3; *P*=.07) or Western Ontario and McMaster Universities Osteoarthritis Index (23.8±17.9 → 21.2±17.4; *P*=.92) at 6 months;<br>• (↑) Hypertension control increased from 60.1 → 76.7% (AEEP/10 keys) and 76.5 → 84.9% (AEEP alone), and diabetes control from 15.0 → 34.9% and 15.5 → 34.1%, respectively.<br>• These community-based programs showed similar improvements in preventive health, mobility, and arthritis outcomes |



XSL•FO
RenderX

| Authors, year | Funding of the program | Strategies/practices/policies of healthy aging | Outcomes |
|---|---|---|---|
| Lee et al [20], 2022 | Research grant to the original RCT[c] | Micro<br>• Integrated geriatric care plus multidomain intervention<br>  • Physical exercise<br>  • Cognitive training<br>  • Diet education<br>  • Chronic condition management | Physio-Cognitive Decline Syndrome<br>• (+) Significant improvement in global cognition (Δ=1.1, 95% CI 0.4 - 1.8; *P*=.003); domain-specific gains in concentration (0.3, 95% CI 0.1 - 0.5; *P*=.01), language (0.2, 95% CI 0.1 - 0.3; *P*=.006), abstraction (0.1, 95% CI 0.0 - 0.3; *P*=.03), and orientation (0.2, 95% CI 0.0 - 0.4; *P*=.01).<br>• (=) No significant change in visuospatial, naming, or delayed memory domains.<br>• ( + ) Improved physical-role-limitation domain of QoL[d] (Δ=5.3, 95% CI 0.3 - 10.4; *P*=.04).<br>• Cognitive impairment, no dementia (CIND)<br>• (–) Significant reduction in frailty (Δ=–0.3, 95% CI –0.5 to –0.1; *P*=.01).<br>• Mobility impairment, no disability (MIND)<br>• (–) Significant reduction in frailty (Δ=–0.3, 95% CI –0.4 to –0.1; *P*=.004) |
| Moriyama et al [21], 2023 | Local Government | Micro<br>• Tai Chi Yuttari exercise performs in sitting or standing position | • (+) Participation in Tai Chi Yuttari classes was associated with significantly longer survival ($\chi^2$=8.782, *P*=.003) and longer long-term care duration ($\chi^2$=5.354, *P*=.02) compared to nonparticipants; effect of survival duration is significant in men only ($\chi^2$=7.875, *P*=.005).<br>• Participation in Tai Chi Yuttari exercise might be effective in delaying death, especially in men, and new certification for long-term care |
| Carter et al [22], 2021 | Research grant | Meso<br>• University unit partners with low-income housing to establish a clinic in the housing estate so that older people can easily access health services, forming an integrated network of care and promoting aging-in-place | After 1 year<br>• (=) Emergency visits and hospitalizations remained stable, but the reasons for emergency visits shifted toward acute issues.<br>• (↑) Follow-up adherence improved slightly, while missed appointments increased a bit.<br>• (+) Use of ancillary services (physiotherapy, occupational therapy, home care, nutritional counseling, behavioral health, etc) expanded significantly.<br>• (↑) Vaccination rates showed a modest improvement overall |
| Lucifora and Villar [23], 2024 | Private company | | |



XSL•FO
RenderX

| Authors, year | Funding of the program | Strategies/practices/policies of healthy aging | Outcomes |
|---|---|---|---|
| | | Micro<br>• Personalized medical health plans<br>  • Physical domain (physical activity courses and Tai Chi classes)<br>  • Mental well-being domain (activity for boosting creativity and memory)<br>  • A multidimensionality of the healthy aging intervention<br><br>Meso<br>• Social domain (activities for improving social engagement) | • (↑) Increased healthy lifestyle adoption by 23% - 25% and social engagement by 18% - 19%<br>• (↓) Reduced poor physical health by 3% - 4% and poor mental well-being by 16% - 17% (in ≥75 years)<br>• (↑) Extended disability-free life expectancy by ≈5 years (14 vs 9 years) |

[a](+)=significant increase/improvement ($P$<.05); (–)=significant reduction ($P$<.05); (↑)/(↓)=nonsignificant trend toward increase/reduction ($P$≥.05); (=)=no significant change/difference.

[b]AEEP: Arthritis Foundation Exercise Program.

[c]RCT: randomized controlled trial.

[d]QoL: quality of life.

Six out of 7 studies involved physical activity in interventions [17-21,23] for healthy aging program. One quasi-experimental study found that Tai Chi participation could significantly increase survival duration especially in male participants [21]. This suggests that demographic characteristics, such as sex, should be explicitly considered when designing and targeting healthy aging interventions. The only cluster RCT in this review showed that preventive interventions addressing hypertension and diabetes control, lifestyle modification, and social contacts could be executed by community health workers, resulting in improvement in mobility, arthritis outcomes, blood pressure, and blood glucose [19]. Another pre-post intervention study using similar preventive strategies for chronic condition management, delivered by local medical staff and community health counselors, also demonstrated improvements in goal attainment for low-density lipoprotein cholesterol, blood pressure control, blood glucose control, and colon cancer screening [18]. These 2 studies illustrate the manpower needed to support health span or healthy aging in communities.

Another illustrative intervention for promoting aging-in-place and minimizing health disparity is the pilot cohort study in America, in which a university unit partnered with a housing operator to set up a primary care clinic in a low-income housing development. One-year follow-up showed that the use of ancillary services (such as physiotherapy, nutritional counseling, etc) increased significantly [22]. This indicates the importance of the removal of transportation barriers to accessing health services in the community.

Three RCT studies designed and implemented a multidomain health promotion education program incorporating components on nutrition [17,20], physical activity [17,20,23], smoking and tobacco control [17], mental well-being [17,23], social communication [17,23], spiritual activities [17], and chronic condition management [20]. These interventions significantly improved active and healthy lifestyles, physical health, and social participation [17,23], reduced physio-cognitive syndrome as well as cognitive and mobility impairment [20], and extended disability-free life expectancy [23]. These findings indicate that, beyond the macro-level strategies, active and healthy aging can be effectively promoted through multidimensional interventions at the micro- and meso-levels, underscoring the need for person- and community-level programs in routine practice. Among these interventions, 57% (4/7) of interventions demonstrated significant improvements in health outcomes. Improvements were effective when they incorporated multidimensional supports, involved participants with preserved functional capacity, and were delivered in settings backed by local community infrastructure.

Four studies reported the development of interventions designed to promote healthy aging (Table 3). All of them were multicomponent, with 1 study covering only components at the micro level (cognition, nutrition, depression, and physical performance) [27], and the other 3 studies intending to address components at the micro- and meso-levels [24-26]. One study emphasized the combination of in-person visits with technology in preparing personalized intervention for physical activity, self-management, social engagement, and digital literacy [24]. Another study focused on a culturally appropriate intervention to promote physical and mental health and spiritual issues at the micro level and social engagement and safety at the meso level [25]. Three studies developed their interventions based on a conceptual framework [24,25,27].



XSL•FO
RenderX

**Table .** Intervention development studies (n=4).

| Authors, year | Framework | Strategies/practices/policies |
|---|---|---|
| Joymangul et al [24], 2024 | Active Ageing and Personalised (AGAPE) | Micro<br>• A supervised learning algorithm combines in-person and technologically personalized intervention on 4 aspects:<br>  • Physical activity<br>  • Digital literacy<br>  • Self-management<br><br>Meso<br>• Social engagement |
| Orellano-Colón et al [25], 2014 | Person Environment Occupational Performance Model | Meso<br>• A culturally appropriate intervention in language, facilitator, metaphors, and content on 6 aspects:<br>  • Activity, Health, and Active Aging (Impact of activity on health, Impact of aging on activities routines)<br>  • Maintaining Physical and Mental Wellbeing through Occupations (Keeping mentally active, Keeping physically active, Nutrition)<br>  • Social Relationships (Building social support through activity participation, Effective communication and social interactions, Coping with loneliness)<br>  • Home and Community Safety (Protection against crimes at home and the community, Fall hazards and home safety, Activity adaptative, assistive devices and home modifications for safety)<br>  • Participation in the Home and Community (Leisure and voluntary activities and health, Transportation use, Finance management, Resources in the community for participation in activities)<br>  • Personal enablers of occupational participants (Role of personal attributes in daily participants, Spirituality as an enabler for occupational participants) |
| Gwyther et al [26], 2018 | —[a] | Micro<br>• Physical activity<br>• Cognitive function<br>• Nutrition<br>• Psychology support<br><br>Meso<br>• Intergeneration activities<br>• Social care |
| Ashikali et al [27], 2024 | WHO's[b] ICOPE[c] approach | Micro<br>• Identify SMART[d] goals based on personal priority regarding:<br>  • Cognitive decline<br>  • Limited mobility<br>  • Hearing loss<br>  • Malnutrition<br>  • Vision impairment<br>  • Depressive symptoms |

[a]Not applicable.

[b]WHO: World Health Organization.



XSL•FO
RenderX

[c]ICOPE: Integrated Care for Older People.

[d]SMART: Specific, Measurable, Attainable, Realistic, and Time-bound.

## Association Studies

We found 5 studies that reported associations of important variables with various measures of healthy aging [28-32]. In these studies, measures were used to conceptualize healthy aging, such as Chinese Healthy Aging Index, Activities of Daily Living or Instrumental Activities of Daily Living, depression, DALYs, and biological age. These studies included 5 variables at the micro level, 2 at the meso level, and 5 at the macro level (Table 4). At the micro level, 4 out of the 5 variables were statistically and significantly associated with measures of healthy aging. Two variables, household expenditure and nutrition, were positively related to measures of healthy aging [28,32]. One study showed alcohol consumption and overwork were positively correlated with DALY, whereas eating habits showed nonsignificance [31]. People with more limitations in Activities of Daily Living were more likely to acquire housing adaptations at Meso level [29]. At the macro level, low housing quality, such as housing poverty, was associated with decreased healthy aging [28], while health insurance coverage and public financing policies were negatively associated with unhealthy ageing [30,31].

**Table .** Association studies (n=5).

| Authors, year | Outcomes | Factors associated with the outcome |
|---|---|---|
| Nie et al [28], 2021 | CHAI[a] (Unhealthy Aging) | Micro<br>• Significant<br>  • (+) Household expenditure<br>Macro<br>• Significant<br>  • (+) Housing poverty |
| Wu and Grundy [29], 2025 | Slower Activities of Daily Living/Instrumental Activities of Daily Living development | Meso<br>• Significant<br>  • (+) Acquiring housing adaptation<br>• Nonsignificant<br>  • Sex |
| Zai [30], 2024 | Poor health; Mobility limitation; Instrumental Activities of Daily Living limitation; Depression | Macro<br>• Significant<br>  • (−) Medical Aging Waiver program expenditure per capita |
| Dragos et al [31], 2022 | Disability-adjusted life years (DALY) | Micro<br>• Significant<br>  • (+) Alcohol consumption<br>  • (+) Overwork<br>• Nonsignificant<br>  • Eating habits<br>Macro<br>• Significant<br>  • (−) Voluntary health insurance<br>  • (−) Quality of institutions<br>  • (−) Public financing (voluntary health insurance) |
| McIntyre et al [32], 2021 | Biological age deceleration | Micro<br>• Significant<br>  • (+) Nutritional components: fiber, magnesium, and vitamin E intake |

[a]CHAI: Chinese Healthy Aging Index.



XSL•FO
RenderX

### Themes From Descriptive Studies

Five studies were descriptive studies that explored participants' perceptions of healthy aging [33-37]. A total of 6 themes at the micro level and 5 at the meso level were identified, but none at the macro level, and they were similar to the interventional components in the interventional, intervention development, and association studies. The 6 micro-level themes included finance, nutrition, physical activity, mental and psychological management, spiritual aspect, and digital literacy, whereas the 5 meso-level themes were independent living/housing, safety, communication, support systems, and health care support (Table 5).

Figure 2 summarizes the key findings of this review.

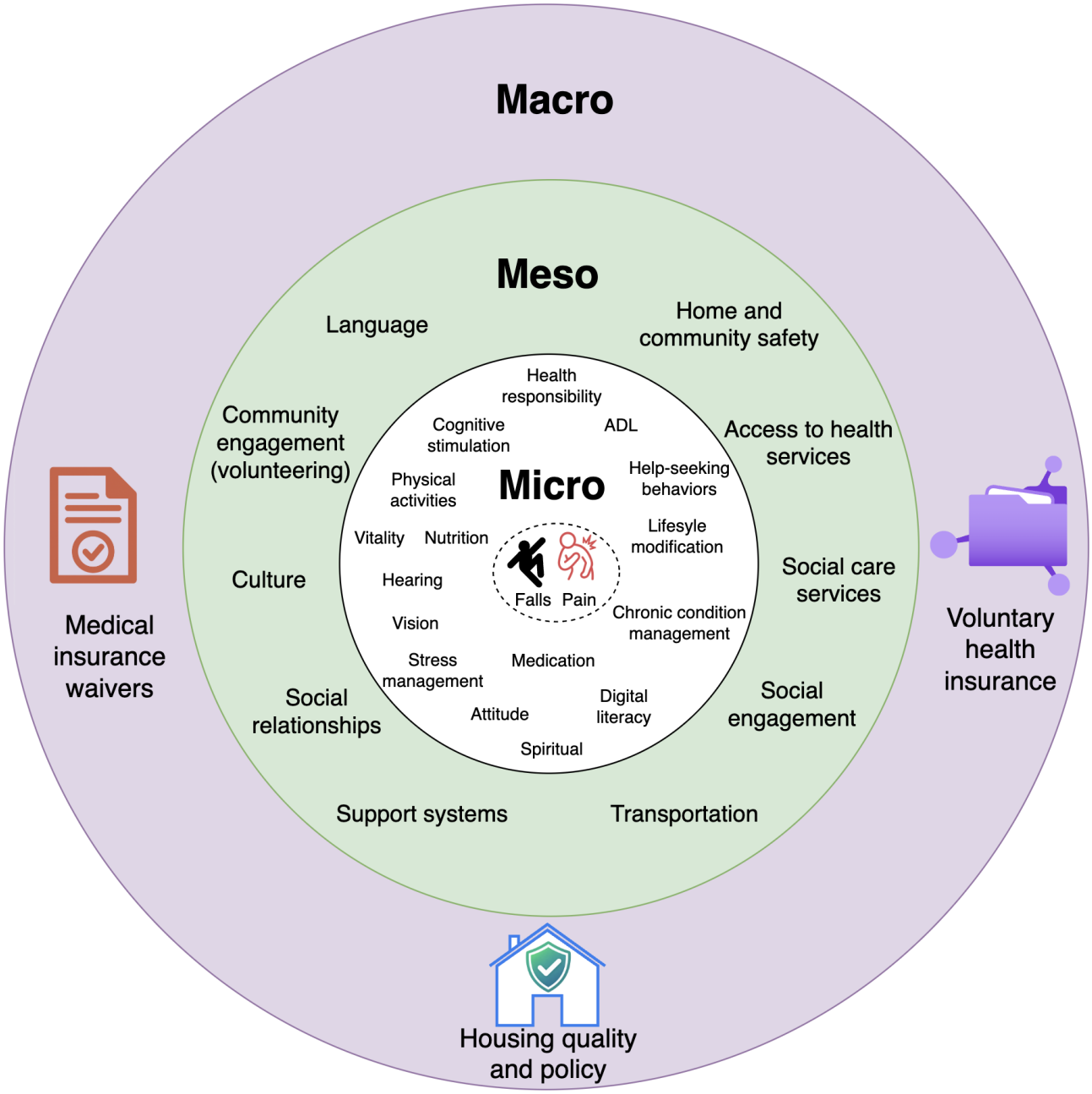
**Table .** Descriptive studies (n=5).

| Micro-, meso-, macro-level framework | Themes identified from the descriptive studies |
|---|---|
| Micro | • Finance [34,37]<br>• Nutrition [34]<br>• Physical activity [34]<br>• Mental and psychological management [34,35]<br>• Spiritual aspect [34]<br>• Digital literacy [33,35] |
| Meso | • Housing and community environment for independent living [33,37]<br>• Safety [33,35,36]<br>• Communication [34,35]<br>• Support system [35,37]<br>• Health care support [36,37] |



XSL•FO
RenderX

**Figure 2.** Strategies, practices, and policies of healthy aging at the micro-, meso-, and macro-levels. ADL: activities of daily living.




## Discussion

### Principal Findings

This scoping review mapped the existing evidence on practices and interventions for healthy aging and health lifespan between 2010 and 2025. We identified 21 studies, of which 15 (71%) were published after 2020 and all studies were conducted in high- and middle-income countries, with 17 (81%) in high-income communities. Four study designs were evident: interventional studies, intervention development studies, association studies, and descriptive studies. Among these practices and interventions, 13 (62%) focused on micro-level factors (eg, physical activity, nutrition, mental and psychological management, spirituality), 11 (52%) focused on meso-level factors (eg, social engagement, support systems, housing and health care support). Very few practices (n=2, 9.5%) addressed macro-level determinants such as public financing or policy. Overall, the findings suggest a growing interest in multidomain programs that integrate micro- and meso-level components, but with important gaps in evidence from low-resource settings and in macro-level strategies.

### Micro- and Meso-Level Factors

Multidomain interventions, including physical exercise, nutrition, emotional management, lifestyle modification, and social engagement, were the most frequently reported interventions that bring benefits to older people to extend health span. This finding is consistent with previous research, showing that interventions can improve physical performance, cognitive function, nutritional status, mood, cardiovascular risk profiles, and overall quality of life at-risk of functional decline [38-40]. From a mechanistic perspective, the rationale for multidomain interventions is also plausible. Aging is driven by multiple



XSL•FO
RenderX

hallmarks, resulting from both endogenous and exogenous stress [41]. Multidomain interventions may act simultaneously on several of these hallmarks [42]. For example, physical activity and nutrition can enhance metabolic regulation, cognitive and social engagement can reduce neuroinflammation, and psychological support can improve stress resilience. Although the included studies rarely measured biological mechanisms directly, this framework provides an explanation for why interventions targeting multiple domains can produce broader benefits of healthy longevity [41].

Nevertheless, multidomain interventions should not be implemented as fixed, one-size-fits-all packages. Their effects may vary according to population characteristics, delivery modes, study settings, intervention intensity, and outcome measures [43,44]. Therefore, how practices and interventions are delivered appears to be important. In this review, evidence across interventional, developmental, associative, and descriptive studies further highlighted the importance of social environmental factors at the meso level, including culture [25], home and community safety [25], housing conditions [29], support systems [35], social care services [26], access to health services [22], and social engagement, etc. Culturally adapted interventions can improve adherence because they are more likely to fit older adults' language, food habits, family structure, religious beliefs, and social expectations [45,46]. Similarly, housing quality and neighborhood safety are important for aging in place. Large-scale analyses have shown that environmental factors account for a substantially greater proportion of variation in mortality risk than genetic predisposition (around 17% vs <2%) [47]. Poor housing may increase exposure and exert particularly strong effects on diseases of the lung, heart, and liver [47], while unsafe homes and communities may increase falls, disability, and hospitalization [48]. In contrast, safe and accessible neighborhoods may support social engagement, mental health, and quality of life [49,50]. Moreover, community-based interventions create opportunities for social interaction to mitigate loneliness and social isolation and thereby influence both psychosocial and biological pathways [51,52]. Given the heterogeneity in intervention components observed in this review, more studies can examine which combinations and dosages of multidomain interventions are most effective for specific populations and delivery conditions.

Our review found that interventions delivered by community health workers improved preventive health behaviors and increased the use of health services [18,22]. These findings indicate that healthy aging interventions also need to be supported by accessible community services and local workforce capacity. Considering both micro- and meso-level factors, future interventions should adopt person-centered designs that match components to older adults' intrinsic capacity, goals, preferences, and social environments, etc. Personalized approaches, such as the ICOPE (Integrated Care for Older People) framework [53], may support the codevelopment of tailored care plans and help translate multidomain interventions into routine practice.

### Macro-Level Factors

At the macro level, long-term care financing policy and health system financing policy, such as medical insurance waiver programs and voluntary health insurance, influence healthy lifespans by reducing financial barriers and improving access to care. Similar to previous multicountry studies, this review found that universal health coverage was positively associated with both life expectancy at birth and healthy life expectancy [54]. Evidence from China further showed that health insurance was associated with higher end-of-life medical expenditure among older adults, suggesting improved access to medical services [55]. However, because all the involved studies were from high-income countries, unequal insurance benefits across different income settings need to be further explored. In addition, none of the interventional and qualitative studies targeted macro-level practices. Future research should move beyond micro- and meso-level approaches and examine how policy, public hospital authority intervention, financing systems, support for low-income housing, and equitable access to services can support healthy longevity. This gap is also important from a life-course perspective. Although most included studies focused on later-life practices and interventions, protection policies are needed to ensure that health-promoting environments and resources are available throughout the life course [56]. Early macro-level interventions aimed at reducing socioeconomic disparities helped narrow health gaps in older age [57]. This is particularly important for older adults living in low-income, rural, minority, or otherwise underserved communities.

### Evidence Gaps in Low-Resource Settings

All studies included in this review were from high- or middle-income countries. While evidence from low-income countries was limited. This limits our understanding of how healthy longevity interventions operate in settings with fewer financial, technological, and service resources. Studies conducted in low-income communities in this review appeared to focus mainly on basic health risk prevention and the use of locally available resources, whereas programs in more affluent settings were more likely to adopt multidomain, service-integrated, or technology-enhanced approaches. This disparity is worthy of attention. Limited financial, health workforce, and technological capacity constrains the feasibility of multidomain interventions in resource-poor settings, which places populations at risk of widening inequalities in access to healthy aging support [37]. Therefore, there is a genuine need for developing and testing low-cost, high-reach strategies for healthy aging, such as community-based, peer-led, and integrated primary care to support older people's functional capacities. Sustained strategies that promote health span require integrated care and collaboration across sectors, including health care professionals, social workers, community organizations and policymakers, particularly for older adults with lower socioeconomic status and multimorbidity. As discussed above, ICOPE is a potential approach [53] to enhance access to care and support intrinsic capacity, but its implementation depends on cross-sector collaboration. Pilot programs in several regions have demonstrated feasibility and acceptability [58]. Future research should draw on implementation science and mixed methods approaches to systematically document the facilitators



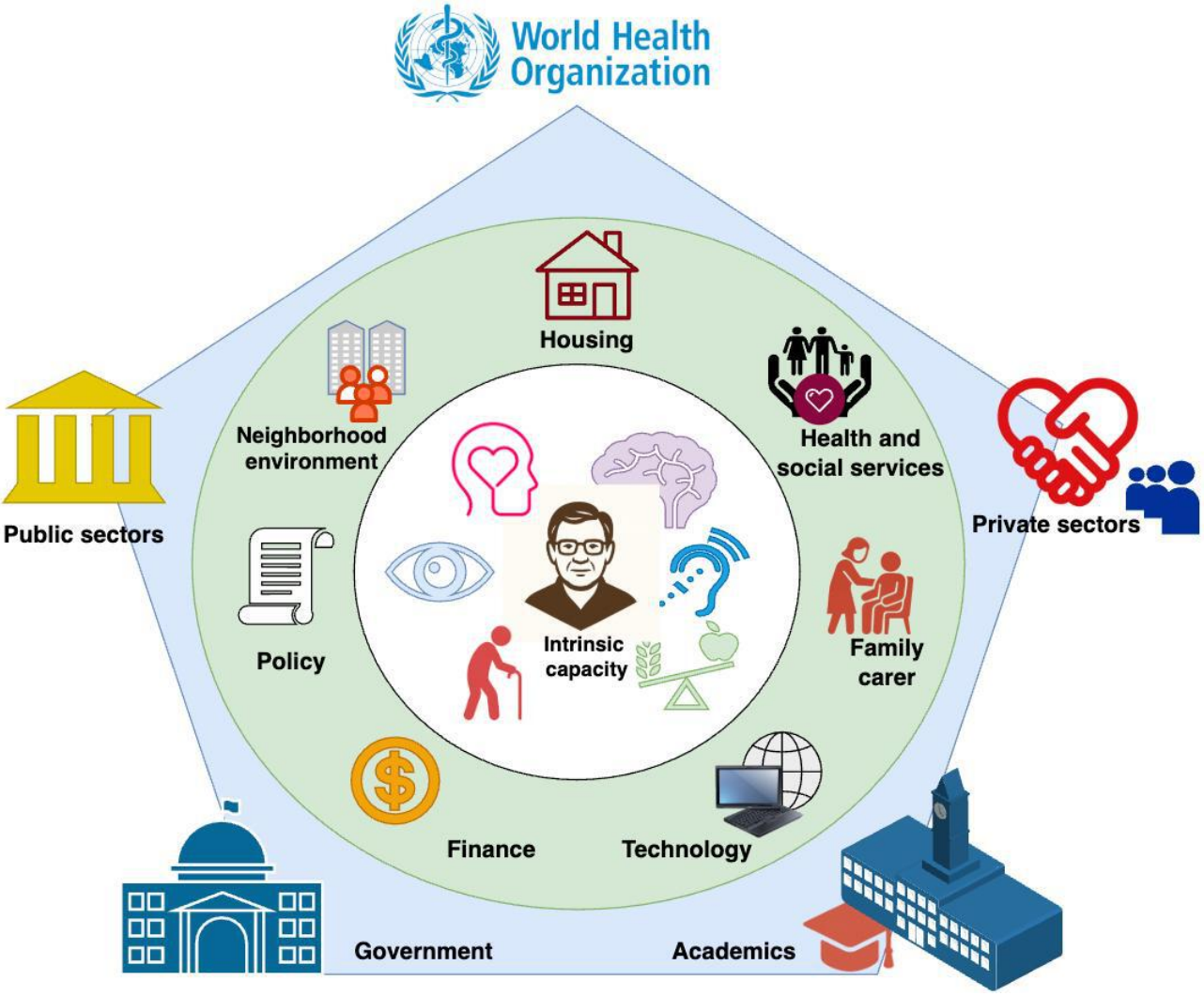
XSL•FO
RenderX

and barriers to intersectoral collaboration, assess readiness and feasibility of related models in diverse health care systems [59]. These steps are crucial to move from small-scale trials toward sustainable system-level practice.

## Future Directions

Based on the findings of this review, we propose a conceptual framework that summarizes how healthy aging and healthy lifespan may be supported through multilevel practices (Figure 3). At the center of the framework is intrinsic capacity, which is shaped by individual-level behaviors and multidomain interventions, meso-level environments and services, and macro-level policies and financing systems. The framework also highlights the importance of cross-sector collaboration among government, public and private sectors, academia, communities, and international organizations.

**Figure 3.** Conceptual framework summarizing multilevel practices supporting healthy aging and healthy lifespan.




## Limitations

This scoping review had a few limitations. First, we focused on studies that explicitly framed their practices or interventions around health span, healthy aging, lifespan, or longevity. So, treatment-oriented interventions such as fall prevention, frailty intervention, or multimorbidity management may have been excluded if they did not explicitly use health span or related terminology, even though such interventions may contribute to healthy longevity. Therefore, the findings should be interpreted as mapping explicitly health span or healthy aging-oriented practices rather than providing a comprehensive review of all interventions that may promote healthy longevity. Second, the components of some initiatives were not well-documented, which may lead us to overlook key elements necessary for extending health span. Third, most included studies (17/21, 81.0%) were conducted in high-income settings, which may limit generalizability. The results should therefore be interpreted with caution, particularly when applying these findings to low-resource contexts. Fourth, the terminological fragmentation observed across studies, where terms such as healthy aging, active aging, successful aging, and healthy life expectancy were used interchangeably. This inconsistent concept may hinder evidence synthesis. Therefore, establishing more standardized terminology would enhance comparability and translation of evidence into practice. Finally, healthy longevity practices or programs that are not documented in peer-reviewed literature were not included. Expanding this review to include gray literature could provide a broader understanding of effective practices in healthy longevity, including those executed by underrepresented populations.



XSL•FO
RenderX

## Conclusion

This scoping review identified key components of community-based strategies, practices, and policies aimed at extending health span and promoting healthy longevity, which warrant further testing and implementation. Overall, the existing evidence suggests that promoting healthy longevity in the community requires cooperation among multiple stakeholders, as well as the provision of integrated, multilevel services or support for older people. Contributions from different sectors and disciplines across all levels are equally important.

Existing interventions remain concentrated in high-income settings and at the micro level, particularly through the use of technology and multidomain integrated services to support older people in living independently and healthily within their communities. In contrast, strategies in low-income contexts largely focus on basic health risk prevention and the mobilization of local resources, such as volunteers, family members, and neighbors to support older people.

Macro-level strategies are especially critical, as policies at this level can promote health equity and reduce socioeconomic inequalities. There is an urgent need to further investigate and develop meso- and macro-level interventions across diverse socioeconomic contexts.

## Acknowledgments

Poe (GPT-5) was used to assist with proofreading and language editing of this manuscript.

## Funding

This project was funded by AIA International (Hong Kong) Limited.

## Authors' Contributions

Conceptualization: AYML, KSLC, JM, IYZ. Funding acquisition: AYML, KSLC. Investigation: AYML, KSLC, JM, IYZ, KLO. Methodology: AYML, KLO, KSLC, JM, IYZ, MQ, KMW, YH. Project administration: KLO, MQ. Resources: AYML, KSLC, JM. Supervision: AYML, JM. Validation: AYML, KLO, JM, IYZ. Visualization: YH, IYZ. Writing – original draft: KLO, MQ, KMW, YH. Writing – review & editing: AYML, JM, IYZ, KSLC

## Conflicts of Interest

Part of the preliminary findings was presented at the Healthy Ageing Conference 2025 in May 2025 in Hong Kong. The funding body had no role in the study design, data collection, analysis, interpretation, or manuscript preparation. To minimize potential bias, the research team conducted the review and analysis independently and was not informed of the funding source during the research process. The authors declare no other conflicts of interest.

Checklist 1
PRISMA-ScR checklist.
[PDF File, 104 KB - i-jmr_v15i1e94085_app1.pdf ]

## References

1. Mather KA, Revelas M. Rethinking ageing research: from lifespan to healthspan. Ageing Longev Res 2025;1(1):6. [doi: 10.53941/alr.2025.100006]
2. Ankrah PK, Mensah ED, Dabie K, Mensah C, Akangbe B, Essuman J. Harnessing genetics to extend lifespan and healthspan: current progress and future directions. Cureus 2024 Mar;16(3):e55495. [doi: 10.7759/cureus.55495] [Medline: 38571872]
3. Garmany A, Terzic A. Global healthspan-lifespan gaps among 183 World Health Organization member states. JAMA Netw Open 2024 Dec 2;7(12):e2450241. [doi: 10.1001/jamanetworkopen.2024.50241] [Medline: 39661386]
4. Venkatapuram S, Amuthavalli Thiyagarajan J. The capability approach and the WHO healthy ageing framework (for the UN Decade of Healthy Ageing). Age Ageing 2023 Oct 28;52(Suppl 4):iv6-iv9. [doi: 10.1093/ageing/afad126] [Medline: 37902511]
5. Scott AJ. The longevity economy. Lancet Healthy Longev 2021 Dec;2(12):e828-e835. [doi: 10.1016/S2666-7568(21)00250-6] [Medline: 36098039]
6. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143. [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
7. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
8. Enhancing healthy longevity: a scoping review of practices and interventions. OSF. URL: https://osf.io/uex2s [accessed 2026-08-01]



XSL•FO
RenderX

9. Economic protocol and title development. In: Joanna Briggs Institute Reviewers' Manual: Joanna Briggs Institute; 2014:88-91 URL: https://dl.icdst.org/pdfs/files3/585eb6f9b7a3929ea6a3c98c6bc9f66d.pdf [accessed 2026-07-30]
10. Healthy ageing and functional ability. World Health Organization. URL: https://www.who.int/news-room/questions-and-answers/item/healthy-ageing-and-functional-ability [accessed 2026-07-30]
11. Crimmins EM. Lifespan and healthspan: past, present, and promise. Gerontologist 2015 Dec;55(6):901-911. [doi: 10.1093/geront/gnv130] [Medline: 26561272]
12. Masfiah S, Kurnialandi A, Meij JJ, Maier AB. Definitions of healthspan: a systematic review. Ageing Res Rev 2025 Sep;111:102806. [doi: 10.1016/j.arr.2025.102806] [Medline: 40523529]
13. EndNote. 2013. URL: https://endnote.com [accessed 2026-08-05]
14. Veritas Health Innovation. Covidence. 2023. URL: https://www.covidence.org/ [accessed 2026-08-05]
15. Serpa S, Ferreira CM. Micro, meso and macro levels of social analysis. Int J Soc Sci Stud 2019;7(3):120. [doi: 10.11114/ijsss.v7i3.4223]
16. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
17. Davodi SR, Zendehtalab H, Zare M, Behnam Vashani H. Effect of health promotion interventions in active aging in the elderly: a randomized controlled trial. Int J Community Based Nurs Midwifery 2023 Jan;11(1):34-43. [doi: 10.30476/IJCBNM.2022.96246.2117] [Medline: 36650846]
18. Newman AB, Bayles CM, Milas CN, et al. The 10 keys to healthy aging: findings from an innovative prevention program in the community. J Aging Health 2010 Aug;22(5):547-566. [doi: 10.1177/0898264310363772] [Medline: 20495156]
19. Zgibor JC, Ye L, Boudreau RM, et al. Community-based healthy aging interventions for older adults with arthritis and multimorbidity. J Community Health 2017 Apr;42(2):390-399. [doi: 10.1007/s10900-016-0268-5] [Medline: 27900515]
20. Lee WJ, Peng LN, Lin MH, Lin CH, Chen LK. Clinical efficacy of multidomain interventions among multimorbid older people stratified by the status of physio-cognitive declines: a secondary analysis from the randomized controlled trial for healthy aging. J Nutr Health Aging 2022;26(10):909-917. [doi: 10.1007/s12603-022-1843-3] [Medline: 36259579]
21. Moriyama N, Kuga T, Oshima T, Sato K, Kurita M, Yasumura S. Association between tai chi yuttari exercise and longevity and prevention of long-term care need: survival analysis in Kitakata City, Japan. Int J Environ Res Public Health 2023 Feb 16;20(4):3472. [doi: 10.3390/ijerph20043472] [Medline: 36834168]
22. Carter DH, Faubert SJ, Henschel E. A novel partnership with low-income housing to support healthy aging. J Nurse Pract 2021 Jun;17(6):718-722. [doi: 10.1016/j.nurpra.2021.01.008]
23. Lucifora C, Villar E. Multi-dimensional healthy aging interventions: evidence from an age-friendly community program in Italy. Ageing Int 2024 Dec;49(4):749-771. [doi: 10.1007/s12126-024-09567-8]
24. Joymangul JS, Ciobanu I, Agnoloni F, et al. Empowering active and healthy ageing: integrating IoT and wearable technologies for personalised interventions. Appl Sci 2024;14(11):4789. [doi: 10.3390/app14114789]
25. Orellano-Colón EM, Varas-Díaz N, Bernal G, Mountain GA. Achieving ecological validity of occupation-based interventions for healthy aging. Phys Occup Ther Geriatr 2014 Dec;32(4):368-380. [doi: 10.3109/02703181.2014.955623] [Medline: 25632170]
26. Gwyther H, Cooke R, Shaw R, Marcucci M, Cano A, Holland C. Perceptions and experiences of frailty interventions: quantitative and qualitative results from a survey of partners within the European Innovation Partnership on Active and Healthy Ageing (EIP-AHA). Ageing Soc 2018;38(9):1843-1867. [doi: 10.1017/S0144686X17000265]
27. Ashikali EM, Ludwig C, Graf C, Ionita I, Busnel C. Application of the Medical Research Council guidance for complex interventions in the development of VIeSA, an intervention to support healthy ageing among community-dwelling older adults. BMC Geriatr 2024 Oct 12;24(1):830. [doi: 10.1186/s12877-024-05399-y] [Medline: 39395966]
28. Nie P, Li Y, Ding L, Sousa-Poza A. Housing poverty and healthy aging in China: evidence from the China Health and Retirement Longitudinal Study. Int J Environ Res Public Health 2021 Sep 21;18(18):9911. [doi: 10.3390/ijerph18189911] [Medline: 34574834]
29. Wu J, Grundy EM. Housing adaptations and older adults' health trajectories by level of initial health: evidence from the English Longitudinal Study of Ageing. Age Ageing 2025 Feb 2;54(2):afaf023. [doi: 10.1093/ageing/afaf023] [Medline: 39957554]
30. Zai X. Evaluating the health outcomes of aging in place: the role of Medicaid Aging Waiver Program on U.S. older adults. BMC Public Health 2024 Aug 5;24(1):2104. [doi: 10.1186/s12889-024-19498-3] [Medline: 39103811]
31. Dragos SL, Mare C, Dragos CM, Muresan GM, Purcel AA. Does voluntary health insurance improve health and longevity? Evidence from European OECD countries. Eur J Health Econ 2022 Nov;23(8):1397-1411. [doi: 10.1007/s10198-022-01439-9] [Medline: 35124741]
32. McIntyre RL, Rahman M, Vanapalli SA, Houtkooper RH, Janssens GE. Biological age prediction from wearable device movement data identifies nutritional and pharmacological interventions for healthy aging. Front Aging 2021;2:708680. [doi: 10.3389/fragi.2021.708680] [Medline: 35822021]
33. Ienca M, Schneble C, Kressig RW, Wangmo T. Digital health interventions for healthy ageing: a qualitative user evaluation and ethical assessment. BMC Geriatr 2021 Jul 2;21(1):412. [doi: 10.1186/s12877-021-02338-z] [Medline: 34215209]




XSL•FO
RenderX

34. Solhi M, Pirouzeh R, Zanjari N. Perspectives on healthy aging in middle age: evidence for health promotion interventions. J Educ Health Promot 2022;11(1):5. [doi: 10.4103/jehp.jehp_972_20] [Medline: 35281398]
35. Puplampu VA, Baffour-Awuah A, Benson PM, Adams L. The impact of older adults' living environment on their healthy aging in place during the COVID-19 pandemic. Perspectives: J Can Gerontol Nurs Assoc 2022;43(1):17-27 [FREE Full text]
36. Somrongthong R, Dullyaperadis S, Wulff AL, Ward PR. The effects of housing on health and health risks in an aging population: a qualitative study in rural Thailand. Biomed Res Int 2014;2014(1):289731. [doi: 10.1155/2014/289731] [Medline: 25101268]
37. Bacsu JR, Jeffery B, Johnson S, Martz D, Novik N, Abonyi S. Healthy aging in place: supporting rural seniors' health needs. Online J Rural Nurs Health Care 2012;12(2):77-87. [doi: 10.14574/ojrnhc.v12i2.52]
38. Chen LK, Hwang AC, Lee WJ, et al. Efficacy of multidomain interventions to improve physical frailty, depression and cognition: data from cluster-randomized controlled trials. J Cachexia Sarcopenia Muscle 2020 Jun;11(3):650-662. [doi: 10.1002/jcsm.12534] [Medline: 32134208]
39. Lee WJ, Peng LN, Lin CH, et al. Effects of incorporating multidomain interventions into integrated primary care on quality of life: a randomised controlled trial. Lancet Healthy Longev 2021 Nov;2(11):e712-e723. [doi: 10.1016/S2666-7568(21)00248-8] [Medline: 36098028]
40. Ngandu T, Lehtisalo J, Solomon A, et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. Lancet 2015 Jun;385(9984):2255-2263. [doi: 10.1016/S0140-6736(15)60461-5] [Medline: 25771249]
41. Guo J, Huang X, Dou L, et al. Aging and aging-related diseases: from molecular mechanisms to interventions and treatments. Signal Transduct Target Ther 2022 Dec 16;7(1):391. [doi: 10.1038/s41392-022-01251-0] [Medline: 36522308]
42. Schneider N, Yvon C. A review of multidomain interventions to support healthy cognitive ageing. J Nutr Health Aging 2013 Mar;17(3):252-257. [doi: 10.1007/s12603-012-0402-8] [Medline: 23459978]
43. Lee WJ, Peng LN, Lin MH, Kim S, Hsiao FY, Chen LK. Enhancing intrinsic capacity and related biomarkers in community-dwelling multimorbid older adults through integrated multidomain interventions: ancillary findings from the Taiwan Integrated Geriatric (TIGER) trial. J Am Med Dir Assoc 2024 May;25(5):757-763.e4. [doi: 10.1016/j.jamda.2023.10.006] [Medline: 37949432]
44. Liang CK, Lee WJ, Hwang AC, et al. Efficacy of multidomain intervention against physio-cognitive decline syndrome: a cluster-randomized trial. Arch Gerontol Geriatr 2021;95:104392. [doi: 10.1016/j.archger.2021.104392] [Medline: 33765656]
45. González-Ballesteros LM, Mojica LE, Riveros S, et al. Evidence-based interventions for promoting healthy aging: a scoping review. J Nutr Health Aging 2026 May;30(5):100834. [doi: 10.1016/j.jnha.2026.100834] [Medline: 41934874]
46. Tziraki-Segal C, De Luca V, Santana S, et al. Creating a culture of health in planning and implementing innovative strategies addressing non-communicable chronic diseases. Front Sociol 2019;4:9. [doi: 10.3389/fsoc.2019.00009] [Medline: 33869336]
47. Argentieri MA, Amin N, Nevado-Holgado AJ, et al. Integrating the environmental and genetic architectures of aging and mortality. Nat Med 2025 Mar;31(3):1016-1025. [doi: 10.1038/s41591-024-03483-9] [Medline: 39972219]
48. Howden-Chapman P, Bennett J, Edwards R, Jacobs D, Nathan K, Ormandy D. Review of the impact of housing quality on inequalities in health and well-being. Annu Rev Public Health 2023 Apr 3;44(1):233-254. [doi: 10.1146/annurev-publhealth-071521-111836] [Medline: 36525958]
49. Stephens C, Bakhshandeh Bavarsad M. Neighborhood qualities are related to better mental health, quality of life, and loneliness over 6 years: pathways through social engagement and social support to aging well. Gerontologist 2025 May 10;65(6):gnaf095. [doi: 10.1093/geront/gnaf095] [Medline: 40045182]
50. Sheth S, Cogle CR. Home modifications for older adults: a systematic review. J Appl Gerontol 2023 May;42(5):1151-1164. [doi: 10.1177/07334648231151669] [Medline: 36655622]
51. Fuchs J, Scheidt-Nave C, Hinrichs T, et al. Indicators for healthy ageing — a debate. Int J Environ Res Public Health 2013 Dec 2;10(12):6630-6644. [doi: 10.3390/ijerph10126630] [Medline: 24317381]
52. Social connection linked to improved health and reduced risk of early death. World Health Organization. URL: https://www.who.int/news/item/30-06-2025-social-connection-linked-to-improved-heath-and-reduced-risk-of-early-death [accessed 2026-07-30]
53. Integrated Care for Older People (ICOPE): guidance for person-centred assessment and pathways in primary care. : World Health Organization; 2025 URL: https://iris.who.int/server/api/core/bitstreams/7516b015-e205-43d1-883f-01f2c31af261/content [accessed 2026-07-30]
54. Ranabhat CL, Atkinson J, Park MB, Kim CB, Jakovljevic M. The influence of universal health coverage on life expectancy at birth (LEAB) and healthy life expectancy (HALE): a multi-country cross-sectional study. Front Pharmacol 2018;9:960. [doi: 10.3389/fphar.2018.00960] [Medline: 30279657]
55. Song L, Zhang X, Wang M. Health insurance and end-of-life healthcare expenditures: evidence from Chinese Longitudinal Healthy Longevity Survey. Npj Health Syst 2026 Apr 27;3(1):26. [doi: 10.1038/s44401-026-00084-1] [Medline: 42527581]
56. Helping people live well at all ages. World Health Organization. URL: https://www.who.int/news/item/26-06-2025-helping-people-live-well-at-all-ages#:~:text=This%20is%20the%20concept%20behind,later%2C%20even%20for%20future%20generations [accessed 2026-07-30]




XSL•FO
RenderX

57. Burton-Jeangros C, Cullati S, Sacker A, Blane D. A Life Course Perspective on Health Trajectories and Transitions: Springer; 2015.
58. Sum G, Lau LK, Jabbar KA, et al. The World Health Organization (WHO) Integrated Care for Older People (ICOPE) framework: a narrative review on its adoption worldwide and lessons learnt. Int J Environ Res Public Health 2022 Dec 22;20(1):154. [doi: 10.3390/ijerph20010154] [Medline: 36612480]
59. Werner A. A Guide to Implementation Research: Urban Institute Press; 2004.

## Abbreviations

**DALY:** disability-adjusted life year
**ICOPE:** Integrated Care for Older People
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
**RCT:** randomized controlled trial
**WHO:** World Health Organization

*Edited by M Balcarras; submitted 24.Feb.2026; peer-reviewed by E Gueven, W Tambo; revised version received 16.Jul.2026; accepted 16.Jul.2026; published 08.Sep.2026.*

*Please cite as:*
*Leung AYM, Cheung KSL, Ou KL, Zhao IY, Qiu M, Wan KM, Montayre J, Huang Y*
*Enhancing Healthy Longevity: Scoping Review of Practices and Interventions*
*Interact J Med Res 2026;15:e94085*
*URL: https://www.i-jmr.org/2026/1/e94085*
*doi:10.2196/94085*

© Angela Yee Man Leung, Karen Siu Lan Cheung, Kai-ling Ou, Ivy Yan Zhao, Min Qiu, Ka Ming Wan, Jed Montayre, Yaqi Huang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 8.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Factors Influencing Adherence to Digital Lifestyle Interventions for Adolescents: Systematic Review and Meta-Analysis of Attrition

Nikolaos Boumparis[1], PhD; Olivia Studhalter[1], MSc; Philippe de Riedmatten[1], MSc; Inci Derya Yücel[1,2], PhD; Kleio Koutra[3], PhD; Katrina Champion[4], PhD; Ana Molina-Barceló[5], PhD; Teresa de Pablo-Pardo[5], MA; Haridimos Kondylakis[6], PhD; Michael Patrick Schaub[1], PhD; Andreas Triantafyllidis[7], PhD; Severin Haug[1], PhD

[1]Swiss Research Institute for Public Health and Addiction, University of Zurich, Konradstrasse 32, Zürich, Switzerland

[2]Institute on Drug Abuse, Toxicology and Pharmaceutical Science, Ege University, Bornova/İzmir, Turkey

[3]Hellenic Mediterranean University Social Applied Research and Social Work Lab, Heraklion, Greece

[4]The Matilda Centre for Research in Mental Health and Substance Use & School of Public Health, University of Sydney, Sydney, Australia

[5]Cancer and Public Health Research Unit, Foundation for the Promotion of Health and Biomedical Research, Valencia, Spain

[6]Computer Science Department, University of Crete & Institute of Computer Science, Heraklion, Greece

[7]Centre for Research and Technology Hellas, Thessaloniki, Greece

**Corresponding Author:**
Nikolaos Boumparis, PhD
Swiss Research Institute for Public Health and Addiction, University of Zurich, Konradstrasse 32, Zürich, Switzerland

## Abstract

**Background:** Intervention studies frequently report that fewer than half of adolescent participants complete digital health interventions as intended, yet a systematic synthesis of the factors driving this low adherence across lifestyle domains is lacking. Understanding these factors is essential for informing the design of interventions that can sustain adherence and achieve meaningful behavioral outcomes.

**Objective:** This systematic review identified and analyzed factors influencing adherence to digital health interventions targeting key adolescent lifestyle domains: physical activity, dietary habits, obesity management, alcohol consumption, and tobacco use.

**Methods:** We systematically searched PubMed, PsycINFO, Embase, and CINAHL in March 2024, with the search updated in June 2026, for studies of digital interventions targeting adolescents (10 - 19 years). Studies were included if they reported adherence outcomes. Data on intervention characteristics, user-related factors, and adherence outcomes were extracted and narratively synthesized using a framework distinguishing intervention-related and user-related factors across 3 evidence types (statistical, descriptive, and qualitative). Study-level attrition was additionally pooled in a random-effects meta-analysis.

**Results:** A total of 116 studies, comprising 46,029 participants, were included. The average completion rate (binary measure) was 66.9% (SD 26%), while average adherence to individual intervention components was 55.2% (SD 25.5%). A meta-analysis of attrition across 108 comparisons yielded a pooled dropout of 16.9% (95% CI 13.6 - 20.9), with very high between-study heterogeneity and no significant moderators. Consistently reported facilitators included personalization, user-friendly design, reminders, goal setting, gamification, and human support (particularly health coaching and parental involvement). The most common barriers were technical difficulties and poor system usability. Evidence regarding the effects of intervention duration and social components on adherence was inconsistent.

**Conclusions:** Given that adolescents frequently underuse digital health interventions, completing only about half of intervention components on average, the effectiveness of these interventions is fundamentally constrained by adherence challenges. Adherence is primarily driven by user-centered design features and robust technical performance, while the impact of social and user-related factors varies by health domain and study context. Future research should adopt standardized adherence frameworks and reporting standards to enable meaningful cross-study comparisons and the development of evidence-based engagement strategies.

***(Interact J Med Res 2026;15:e84822)*** doi:10.2196/84822

**KEYWORDS**

adherence; digital health interventions; adolescents; health promotion; systematic review



XSL•FO
RenderX

## Introduction

Digital health interventions, including web-based platforms, mobile apps, wearable device interfaces, and technology-integrated school programs, are increasingly used to promote behavioral change among adolescents. The proliferation of these interventions is driven by several factors: adolescents' high affinity for digital technologies, the scalability and cost-effectiveness of digital delivery models, and the growing recognition that health-promoting behaviors established during adolescence serve as foundational determinants for preventing noncommunicable diseases across the lifespan [1,2]. Recent meta-analytic evidence, however, suggests that the effects of digital health interventions on adolescent lifestyle behaviors are small and often short-lived, with limited sustained impact on outcomes such as physical activity, dietary behavior, and substance use [3]. Despite these modest effect sizes, the development and deployment of digital health interventions for adolescents continues to accelerate, propelled by commercial interest, low-cost dissemination through social media channels, and the accessibility of app-based delivery for a population that increasingly turns to digital solutions before seeking in-person support [4]. This trajectory makes it important to understand why these interventions often fail to achieve their intended outcomes.

A key factor underlying the limited effectiveness of digital health interventions is suboptimal adherence. Research consistently shows that many users discontinue digital health apps prematurely, directly compromising intervention efficacy and therapeutic outcomes [5,6]. For the purposes of this review, we adopt the conceptual framework proposed by Perski et al [7], which distinguishes between behavioral engagement (the extent of usage, ie, the amount of the intervention that is used) and experiential engagement (the subjective experience of using the intervention, eg, interest, attention, and affect). We focus specifically on the behavioral dimension of engagement, hereafter referred to as adherence, defined as the extent to which adolescents engage with digital interventions as intended by the developers, including frequency of use, completion of prescribed activities, and sustained participation over the intended duration. We do not systematically assess subjective engagement experiences (eg, satisfaction and perceived helpfulness), though these are noted where reported in included studies. This focus aligns with established frameworks for measuring engagement in digital behavior change interventions [8,9]. The challenge of sustaining adherence is particularly pronounced among adolescents, who may require engagement strategies tailored to their developmental stage to maintain consistent participation [4].

Adherence to digital health interventions has been extensively studied in adult populations, particularly in the context of mental health and chronic disease management, where systematic investigations have identified key predictors including user motivation, intervention design characteristics, and technology usability factors [5,10]. However, adolescents differ from adults in developmentally significant ways that may affect their engagement with digital technologies. Adolescents are navigating a critical period of identity formation and social development, during which cognitive functions including sustained attention, impulse control, and decision-making capacity are still maturing [11,12]. These developmental characteristics suggest that determinants of adherence identified in adult populations may not directly apply to adolescent users, necessitating a population-specific evidence synthesis.

Several categories of factors have been proposed as influencing adherence to digital health interventions among adolescents. At the individual level, intrinsic motivation, self-efficacy beliefs, and developmental readiness for behavioral change have been identified as potential determinants, while household-level factors such as parenting practices, family support structures, and socioeconomic considerations may also play a role [13,14]. At the intervention level, technical design elements including personalization algorithms, automated reminder systems, intuitive user interfaces, goal-setting mechanisms, and self-monitoring capabilities have been proposed as critical determinants of sustained engagement [15,16]. Various strategies to enhance engagement have also been suggested, including participatory co-design approaches, adaptive personalization protocols, and just-in-time intervention delivery [4]. However, the empirical evidence for these factors has not been comprehensively synthesized for adolescent populations specifically.

While previous systematic reviews and meta-analyses have examined the overall effectiveness of digital health interventions for adolescents [2,17] and investigated determinants of adherence in mobile health apps among general adult populations [5,18], no comprehensive synthesis exists that specifically focuses on factors influencing adherence within digital health interventions targeting lifestyle behaviors among adolescents. This represents a significant knowledge gap. Without such a synthesis, researchers and developers lack an evidence base to inform the design of interventions that can sustain adolescent adherence, which is a prerequisite for achieving meaningful behavioral outcomes. Given that average adherence rates in digital health interventions are around 50%, even modest improvements in sustained adherence could substantially enhance intervention effectiveness at the population level.

This systematic review therefore aims to identify, categorize, and synthesize the empirical evidence on factors that influence adherence to digital lifestyle interventions among adolescents aged 10 - 19 years. By examining evidence across multiple health domains, namely physical activity, dietary habits, obesity management, alcohol consumption, and tobacco use, we provide a cross-domain synthesis that reveals both consistent determinants and domain-specific patterns. The findings are intended to generate actionable insights for researchers, intervention developers, and policymakers working to design digital health interventions that achieve sustained adherence among adolescent populations.



XSL•FO
RenderX

## Methods

### Identification of Studies

The review was guided by the following Population-Exposure-Outcome (PEO) framework: Population: adolescents aged 10‐19 years; Exposure: digital health interventions (web-based, app-based, or text message–based) targeting one or more of 5 lifestyle health domains (physical activity, dietary habits, obesity management, alcohol consumption, and tobacco use); Outcome: adherence to the digital intervention, operationalized as the extent of use relative to intended use parameters specified by intervention developers, including completion rates and component-level adherence metrics.

A comprehensive literature search was conducted across 4 electronic databases: PubMed, PsycINFO, CINAHL, and Embase in March 2024. No restrictions were applied regarding publication date, language, or geographic region. The search strategy was developed iteratively by one researcher (NB) and used combinations of controlled vocabulary terms (MeSH, Emtree, CINAHL Subject Headings, and PsycINFO Thesaurus terms) and free-text keywords across 3 key concepts: adherence or engagement, digital health interventions, and specific health domains (physical activity, diet, obesity, alcohol, and tobacco). While our inclusion criteria encompassed internet-, computer-, and smartphone-based interventions, the term "digital interventions" is used throughout this review for the sake of brevity. The initial screening process involved reviewing titles and abstracts, followed by a detailed assessment of full-text articles for studies that potentially met our inclusion criteria. Two researchers from our team independently performed the search and screening processes, resolving any discrepancies through discussion until consensus was reached (OS and PdR). Discrepancies at each stage were resolved through discussion between the two reviewers. In cases where consensus could not be reached, the senior members of our team (NB and SH) were consulted. The complete search strategy, including detailed search strings, can be found in Multimedia Appendix 1. The search was updated on June 22, 2026, to capture studies published since the original search. Using the identical search strings across the same 4 databases (PubMed, PsycINFO, CINAHL, and Embase), we retrieved records covering the period from March 2024 to June 2026. The deduplicated updated records were screened against the original eligibility criteria; title and abstract screening, full-text screening, and data extraction of the newly included studies were performed by 2 researchers (IDY and NB), with discrepancies resolved through discussion. This systematic review was conducted and reported in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines [19].

### Eligibility Criteria

Studies examining digital interventions targeting adolescents aged 10‐19 years (as defined by the World Health Organization [WHO]) were required to focus on at least one of the following 5 specified health domains: physical activity, diet, obesity management, alcohol use, and tobacco use. Studies including some participants older than 19 years were retained when the intervention primarily targeted adolescents or youth and the sample mean age was within or very close to the 10‐to 19-year adolescent age range. These 5 health domains were selected based on their established role as modifiable lifestyle factors associated with the prevention of cancer. The selection was informed by the SUNRISE project (Sustainable Interventions and Healthy Behaviours for Adolescent Primary Prevention of Cancer with Digital Tools; Horizon Europe, Grant Agreement Number 101136829), within which this review was conducted.

These domains were selected based on their established links with chronic disease (eg, cancer, cardiovascular disease, diabetes, and so on) prevention and represent modifiable lifestyle factors that have been shown to reduce cancer risk throughout life. Eligible studies were required to provide qualitative or quantitative data on the intended use parameters or actual adherence patterns with digital intervention components as defined by the developers.

No restrictions were applied regarding publication date or language. Interventions had to implement digital health promotion strategies as the primary therapeutic component; studies in which digital strategies merely supplemented in-person interventions were deemed ineligible. Intended use was operationally defined as the minimum level of digital platform interaction specified by the intervention developers as necessary for achieving therapeutic adherence. Actual use, on the other hand, encompassed the adherence patterns that were empirically measured within the study population. Studies combining multiple intervention modalities, such as automated messaging platforms integrated with counseling calls, were included, provided adherence measures could be isolated and attributed specifically to digital components. Supplementary elements were categorized as "personal support."

To focus on primary empirical evidence, we excluded systematic reviews and meta-analyses. Single-session interventions were excluded as they preclude meaningful longitudinal adherence assessments, resulting in universal participation rates that provide limited insight into sustained adherence patterns. Studies that used digital platforms solely for communication (eg, email-based coaching or video consultation sessions) were excluded as they represent interpersonal rather than digital behavioral intervention modalities. Additionally, studies that primarily targeted families or other populations rather than adolescents directly were deemed to be outside the scope of this review.

### Data Extraction

Data from studies included in the initial systematic literature search were extracted independently by 2 researchers (OS and PdR), while data for studies identified through the updated systematic search were extracted independently by 2 researchers (IDY and NB). Any discrepancies were resolved through discussion. The following information was extracted for each included study: general study characteristics, study population characteristics, intervention characteristics, and information on intervention use.

General study characteristics included the country, study design, and health domain. Study population characteristics comprised



XSL•FO
RenderX

mean age, gender distribution (in %), target group (high school students, university students, and others such as adolescents not enrolled in formal education, those in vocational training programs, community-based youth groups, or clinical populations), intervention setting (online, school-based, home-based, community-based, or clinic-based), presence of clinical diagnosis, number of participants, and inclusion and exclusion criteria. Intervention characteristics included the intervention name, delivery type (web-based, through an app, and through text messages), presence and type of personal support, and provision of participant incentives. Information on intervention use encompassed study duration, number of modules, intended use, actual participant adherence, and attrition rate. Additionally, we extracted data on intervention effectiveness and all reported factors influencing adherence. These factors were classified as either intervention-related, that is, modifiable elements of the intervention design, or user-related, that is, inherent characteristics of the participants. The synthesis of studies identified through the initial systematic search was conducted by OS and reviewed by PdR, whereas the synthesis of studies identified through the updated systematic search was conducted by IDY and reviewed by NB. Factor categorizations and evidence classifications were cross-checked against the original study manuscripts. Any disagreements were resolved through discussion with the senior team members (NB and SH).

### Data Synthesis

To guide our synthesis, we adopted a framework informed by the conceptualization of engagement with digital behavior change interventions proposed by Perski et al [7], which distinguishes between behavioral engagement (extent of usage) and experiential engagement (subjective experience). This review focuses on behavioral engagement, operationalized as adherence: the degree to which participants used the intervention as intended by developers. Adherence was captured through two primary metrics where reported: (1) completion rate, defined as the proportion of participants who completed the entire intervention program (binary measure), and (2) component adherence, defined as the mean proportion of individual intervention components completed by participants (continuous measure). The factors influencing adherence were categorized as intervention-related (modifiable elements of intervention design and delivery) or user-related (participant characteristics), and further classified by evidence type: statistical (inferential testing, $P<.05$), descriptive (quantitative patterns without significance testing), or qualitative (interviews, focus groups, and open-ended responses).

As a first step, the identified studies were categorized according to their respective health domains. In the second step, factors influencing adherence reported in the reviewed articles were qualitatively evaluated and summarized in individual tables for each health domain (Tables S1-S6 in Multimedia Appendix 1 [20-110]). These factors were then categorized into intervention-related factors that acted as either facilitators or barriers to adherence, as well as user-related factors that either facilitated or hindered adolescents' adherence to digital interventions. In the third step, the factors were analyzed based on 3 types of evidence: statistical significance (findings derived from inferential statistical analyses with $P<.05$), descriptive evidence (quantitative patterns identified through descriptive statistics without formal significance testing), and qualitative evidence (findings from interviews, focus groups, and open-ended survey responses). The compiled tables present an integrated view of these 3 evidence types to provide a comprehensive understanding of factors influencing adherence.

To complement the narrative synthesis, we conducted a quantitative synthesis of attrition. For each intervention arm reporting sufficient data, we calculated the proportion of enrolled participants lost during the intervention and meta-analyzed the logit-transformed proportions using a random-effects model, with the between-study variance estimated using restricted maximum likelihood and CIs calculated using the Knapp-Hartung adjustment [111]. Summary estimates and CIs were backtransformed to the proportion scale. Heterogeneity was quantified using $\tau^2$, $I^2$, and a 95% prediction interval. Robustness was examined in sensitivity analyses using a binomial-normal generalized linear mixed model with a logit link [112,113], and a 3-level random-effects model in which estimates from multiple intervention arms were nested within studies [114]. Subgroup analyses and random-effects meta-regression examined health domain, delivery modality, level of personal support, provision of incentives, intervention duration, and mean participant age as potential moderators. Potential small-study effects were explored using funnel plots and an Egger-type regression test [115] because conventional funnel-plot methods may be misleading in meta-analyses of proportions; these analyses were considered exploratory. Reported adherence determinants were not quantitatively pooled because of substantial heterogeneity in their measurement. Instead, we summarized the direction and consistency of each determinant across health domains and evidence types and reported this synthesis in accordance with the Synthesis Without Meta-analysis (SWiM) reporting guideline [116]. Analyses were conducted in R (version 4.5.3; R Core Team) [117] using the *metafor* package [118].

## Results

### Selection and Inclusion of Studies

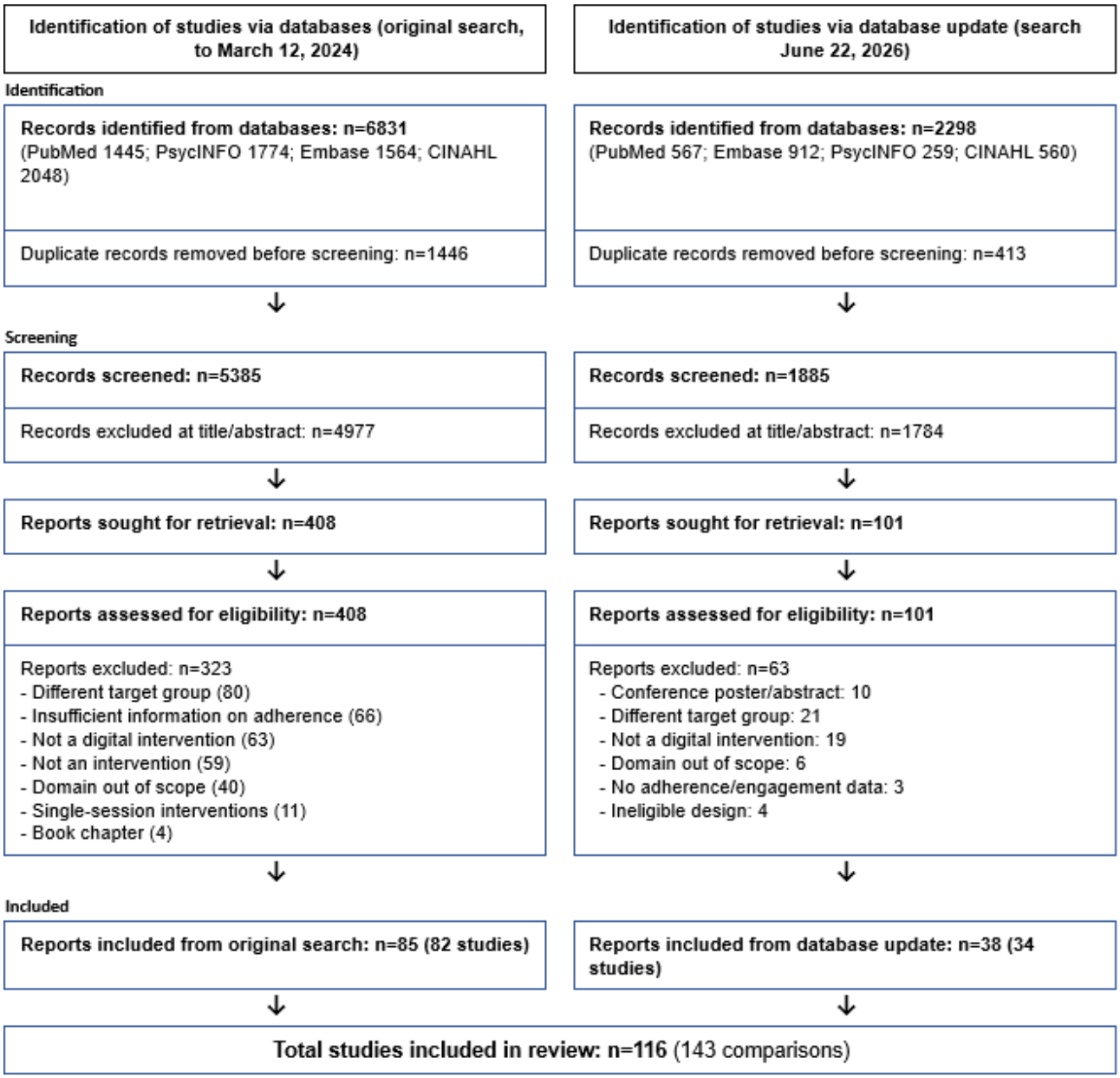
The original search of the electronic databases was performed on March 12, 2024, and yielded 6831 records. After removal of duplicates, 5385 records remained for title and abstract screening, and the full texts of 408 articles were subsequently examined, of which 85 records (82 studies) [20-76,119-146] were included. The search was updated on June 22, 2026, across the same 4 databases and identified 2298 records; after removal of 413 duplicates, 1885 records were screened by title and abstract, of which 101 were sought for retrieval and assessed as full texts. Of these, 63 were excluded (10 conference posters or abstracts, 21 for an ineligible population, 19 because the digital component was not the primary intervention, 6 for a health domain outside the review scope, 3 for reporting no adherence or engagement data, and 4 for an ineligible study design), leaving 38 newly included records. These corresponded to 34 additional study entries after publications reporting on the same participant cohort were combined, yielding a total of 116



XSL•FO
RenderX

included studies [20-69,71-109,119-144,147]. Figure 1 visualizes the selection process and reasons for exclusion for both the original and updated searches.

**Figure 1.** Flowchart for inclusion of studies.




## Characteristics of Included Studies

The 116 included studies comprised several designs: 77 (66.4%) were randomized controlled trials (RCTs) [2,6,8,9,11,14,16,18,19,21,23,24,28,29,31,33,34,36,38,41,43,44,46,48,49,51,53,55,57,58,60,62,63,66,68,69,71,73,74,76,77,79,81,83,87,89,90,93,94,96,98,100,101,103,105,106,109,119,120,123,124,127,128,130,131,133,135,137,138,140,141,143,144,147] 36 (31%) were quasi-experimental studies [20,23,25,27,29,30,32,34,36,42,44,45,47,50,52,61,62,64,65,67,78,80,84,88,91,99,103,112,122,130,141,142], and 3 (2.6%) were observational studies [56,86,95].

Geographically, the included studies spanned 26 countries across 5 continents, with persistent regional disparities. The United States remained the single largest contributor, accounting for 43.1% (50/116) studies, followed by Canada, Spain, and Australia. European studies again formed a substantial group, and the updated search broadened the geographic spread by adding studies from countries not previously represented, including Brazil, Uzbekistan, Vietnam, Taiwan, India, Turkey, and Portugal.

The analysis included 143 comparisons from the 116 included studies, as 23 studies [22,26,35,39,40,52,57,67,81,82,85,87,89,91,97,104,107,108,126,129,134,136,147] contained more than one intervention group relevant to our systematic review (21 contributed 2 intervention arms and 2 [85,108] contributed 4 intervention arms). Studies published on the same participant cohort were combined into single entries in the tables and handled as one study. The median intervention duration was 84 (IQR 42-135; range 2-730) days, with 27/143 (18.9%) comparisons lasting 1 - 30 days, 54/143 (37.8%) lasting 1 to 3 months, 44/143 (30.8%) lasting 3 to 12 months, and 4/143 (2.8%) lasting 1 year or longer. Fourteen comparisons out of 143 comparisons (9.8%) did not report intervention duration. Of the 143 comparisons, 79 (55.2%) provided no incentives to



XSL•FO
RenderX

participants, 44 (30.8%) offered monetary or material rewards such as cash, gift cards, vouchers, prize draws, or a retained device, and 20 (14%) offered nonmonetary incentives such as academic credit or gamified points. A detailed description of the study characteristics can be seen in Tables S7-S18 in Multimedia Appendix 1 [20-110,119-144,147].

## Characteristics of Study Populations

The included studies comprised a total of 46,029 participants summed across intervention and control arms, with a median of 52.5 (IQR 28-232; range 10-5260) participants per study. Of the 142 comparisons that reported exact participant numbers, 91 (64.1%) included fewer than 100 participants, 33 (23.2%) had between 100 and 500 participants, 9 (6.3%) included 500 to 1000 participants, and 9 (6.3%) had more than 1000 participants. The overall mean age was 14.48 (SD 2.07) years. Across the studies reporting participant sex, female participants comprised 59.1% of the combined sample (115 [20-69,71-87,90-109,119-144,147] of 116 studies reported sex), essentially unchanged from the original cohort's 59.2.

The great majority of studies enrolled adolescents with no prior clinical diagnosis. Among studies recruiting clinically defined populations, obesity or overweight remained the most common, followed by childhood cancer, congenital or coronary heart disease, type 1 diabetes mellitus, and juvenile idiopathic arthritis. The update introduced clinical populations absent from the original synthesis, including single studies of chronic kidney disease, cystic fibrosis, elevated blood pressure or cardiovascular risk, and adolescents recruited through mental health services, together with a larger group of studies enrolling adolescents who used nicotine or electronic cigarettes.

## Characteristics of Digital Interventions

Across the 143 reviewed comparisons, 71 (49.7%) were delivered through location-independent digital platforms accessible via smartphones or personal devices, 33 (23.1%) were school-based, 19 (13.3%) were home-based, 11 (7.7%) were clinic-based, 5 (3.5%) were community-based, and 4 (2.8%) combined settings such as summer camp plus online or school plus mobile delivery.

Regarding delivery mode, 51 (35.7%) comparisons were app-based and 50 (35%) were internet- or web-based, together accounting for roughly 7 in 10 interventions; 13 (9.1%) paired a wearable activity tracker or accelerometer with a smartphone or web platform, 11 (7.7%) relied exclusively on text messaging, 9 (6.3%) were computer- or software-based, 8 (5.6%) used a mixed multimodal approach, and 1 (0.7%) delivered exercise through immersive virtual reality.

Personal support beyond initial recruitment was provided in 72 comparisons (50.3%), most commonly through research-team contact, health-professional input, or coaching, while the remaining 71 (49.7%) offered no personal support or relied solely on automated prompts and reminders.

Effectiveness was determined based on whether studies demonstrated statistically significant improvements in their prespecified primary outcomes, distinct from adherence metrics; 52.8% (47/89) of the interventions evaluated for effectiveness were reported as effective. The mean intervention completion rate across the 68 comparisons reporting it was 66.9% (SD 26%), and component adherence averaged 55.2% (SD 25.5%) across the 66 comparisons that reported it, with the remaining comparisons not reporting full completion or quantitative adherence data.

## Characteristics of Health Domains

As displayed in Tables S7-S18 in Multimedia Appendix 1 [20-110,119-144,147], the included studies were categorized into 5 distinct health domains based on their intervention focus (physical activity, diet, obesity, alcohol use, and tobacco use), with an additional category for multicomponent lifestyle interventions that targeted multiple health domains. Of the 116 included studies [20-69,71-109,119-144,147], 31 (26.7%) [61-69,71-76,98-109,141-143,147] were multicomponent lifestyle interventions, 32 (27.6%) [20-33,77-86,119-125,144] focused on physical activity, 19 (16.4%) [34-48,127-130] targeted obesity, 14 (12.1%) [55-57,90-94,131-136] targeted tobacco use, 10 (8.6%) [49-54,87-89,126] focused on diet, and 10 (8.6%) [58-60,95-97,137-140] addressed alcohol use.

Within the 31 multicomponent lifestyle interventions, the most common combination was diet and physical activity (12 studies, 38.7% [61,62,64-66,68,74,76,102,104,107,142]), followed by alcohol and tobacco (5, 16.1% [63,69,71,72,143]), diet and obesity (4, 12.9% [100,101,105,147]), all 4 of alcohol, diet, physical activity, and tobacco simultaneously (4, 12.9% [73,75,103,109]), diet, physical activity and obesity (3, 9.7% [67,98,99]), physical activity and obesity (2, 6.5% [106,108]), and diet, physical activity, and tobacco (1, 3.2% [141]).

Across the health domains, dietary interventions showed the highest mean intervention-completion rate (83.9%, SD 21%) and alcohol-focused interventions the lowest (48.4%, SD 35.4%); for component adherence, dietary interventions again showed the highest mean rate (63.7%, SD 23%, essentially level with physical activity interventions at 62.8%), while obesity interventions showed the lowest (45.9%, SD 22.5%). These domain-level figures rest on small and unevenly distributed numbers of reporting comparisons (n=4 - 24 per domain) and on widely overlapping ranges, and are therefore indicative rather than definitive. In the meta-analysis of attrition, pooled dropout was lowest for physical activity and highest for alcohol interventions, as detailed in the "Quantitative Synthesis of Attrition" section.

The findings presented below are organized according to 3 distinct levels of evidence synthesis: statistical evidence, descriptive evidence, and qualitative evidence. Where individual studies did not generate all 3 types of evidence, only the categories of evidence that were empirically identified are reported and analyzed in the subsequent sections. For a detailed description of all the reported quantitative adherence values, please refer to Tables S8, S10, S12, S14, S16, and S18 in Multimedia Appendix 1 [20-110,119-144,147].

## Quantitative Synthesis of Attrition

Across the included studies, 108 comparisons (88 studies [illegible] 30,956 participants) reported extractable attrition data. The

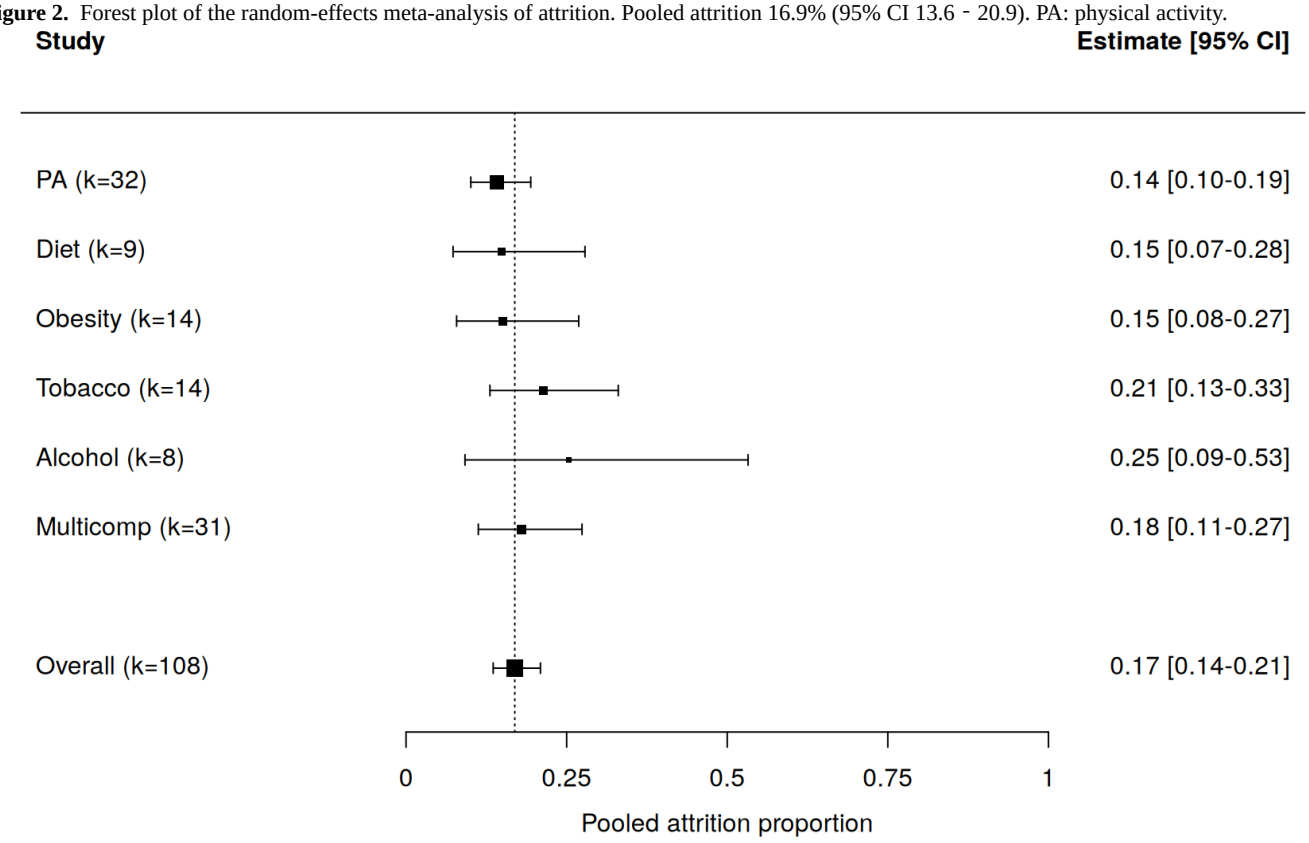


XSL•FO
RenderX

pooled attrition proportion was 16.9% (95% CI 13.6-20.9). Heterogeneity was very high ($I^2$=98.3%; $\tau^2$=1.51), and the 95% prediction interval was correspondingly wide (1.7-70.3), indicating that attrition varied substantially across study contexts. Estimates were robust to model choice (exact binomial-normal model: 14.3%, 95% CI 11.1-18.2; 3-level model accounting for the nesting of multiple intervention arms within studies: 17.3%, 95% CI 13.6-21.7).

Pooled attrition was lowest for physical activity (14.1%, 95% CI 9.9-19.7; k=32), diet (14.9%, 95% CI 6.2- 31.5; k=9), and obesity (15.1%, 95% CI 7.3-28.4; k=14) interventions, intermediate for multicomponent interventions (17.9%, 95% CI 11.1-27.8; k=31), and highest for alcohol (25.3%, 95% CI 7.0-60.5; k=8) and tobacco (21.4%, 95% CI 11.9-35.5; k=14) interventions, although the CIs overlapped substantially (Figure 2). None of the examined moderators reached significance in the omnibus Wald-type tests of the meta-regression models (QM): health domain ($QM_5$=0.79, *P*=.56), delivery modality ($QM_3$=1.22, *P*=.30), personal support ($QM_1$=0.57, *P*=.45), incentives ($QM_1$=1.03, *P*=.31), intervention duration ($QM_1$=2.60, *P*=.11), or mean participant age ($QM_1$=0.62, *P*=.43), with each moderator explaining essentially none of the between-study variance. An Egger-type regression test indicated possible small-study effects (z=−2.75, *P*=.007), with smaller studies tending to report lower attrition, although this test should be interpreted cautiously at this degree of heterogeneity (Figure 3).

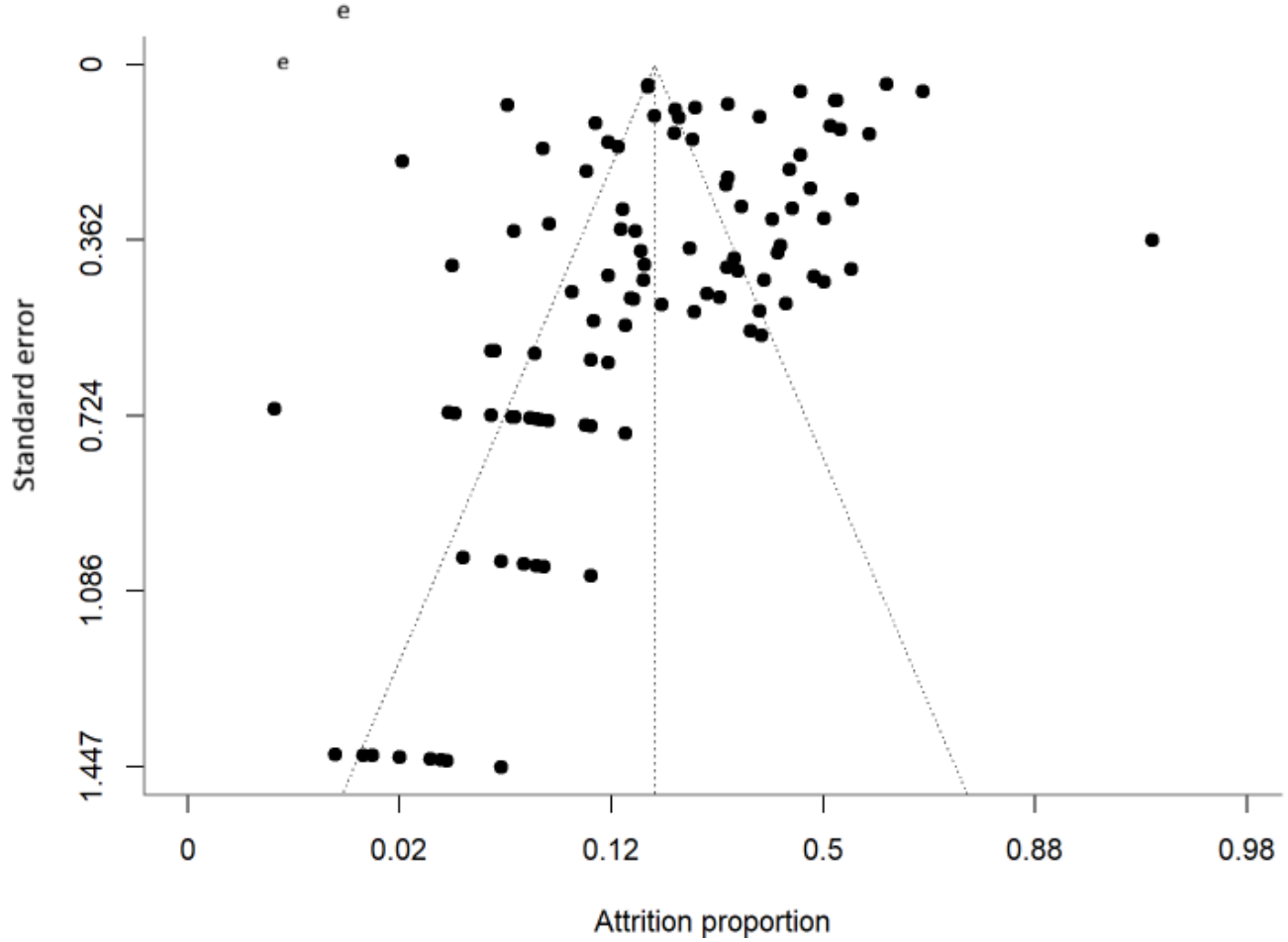
**Figure 2.** Forest plot of the random-effects meta-analysis of attrition. Pooled attrition 16.9% (95% CI 13.6 - 20.9). PA: physical activity.






XSL•FO
RenderX

**Figure 3.** Funnel plot for the attrition meta-analysis (Egger-type test z=–2.75, *P*=.007).




Notably, pooled attrition (approximately 17%) was considerably lower than the degree of underuse implied by the completion and component-adherence metrics, under which roughly one-third of participants did not complete the full intervention, and approximately 45% of intervention components went unused. This pattern indicates that many adolescents remained nominally enrolled while substantially under-using the intervention, so that retention statistics alone overstate true adherence. Overall, attrition in adolescent digital lifestyle interventions is substantial but highly heterogeneous and is not explained by standard study-level characteristics.

## Physical Activity

### *Intervention-Related Factors*

Several intervention design features were associated with adherence to physical activity interventions. Statistical evidence indicated higher adherence during daytime hours [20] and on weekdays versus weekends [21], while longer study duration was negatively associated with adherence [20,22]. Descriptive evidence reinforced the importance of specific program features, with activity trackers [23-25], outdoor and solitary usage [26], and targeted components including workout missions, storylines, parent-child activities, physical activity education, run logging, and week-runs [26] all showing positive associations with adherence. Further descriptive facilitators included the “Kids Zone” interface [24], coaching sessions [25], and coach reminder messages [21]. Conversely, descriptive evidence identified technical difficulties and equipment dissatisfaction [23,27], social intervention features [26], weekly challenges and Facebook videos (Meta Platforms, Inc) [28], poorly timed text messages during class time [28], educational modules [25], and excessive video duration [21] as barriers.

Qualitative evidence provided the richest insights into design-related facilitators and barriers. Participants consistently endorsed goal-setting components [28,29], tracking and monitoring features [26,28], flexible goal reminders [30], clear instructions with progressive difficulty scaling, encouragement, and immersive app design [26], and engaging activities including gamification elements and peer competition [27,28]. The provision of activity trackers (eg, Fitbits) was viewed positively [24], as were activity logging and local resource information [29], incentives [31], and diverse content including varied video and sports options [20,22]. However, qualitative data also revealed substantial technical barriers, including synchronization issues, device charging logistics, and website login difficulties [24,28,29,31,32], as well as equipment dissatisfaction due to perceived device inaccuracy [27,28,30]. Participants reported competition from alternative systems and information sources [24,28], unappealing website design with excessive text and insufficient youth-oriented content [24,29], limited social components [29], inadequate reward systems [31], and specific features such as message boards and expert forums that reduced adherence [29]. Participants expressed preferences for text



XSL•FO
RenderX

messages over emails, which often went unread due to spam filtering [32].

The updated search added 10 physical activity studies that broadly reinforced these design-related patterns while adding nuance around support and gamification. Personalization again emerged as a facilitator, with tailored nurse feedback [77] and readiness-based content routing [78] supporting engagement, whereas a lack of customization lowered perceived usability [79]. Human support in its various forms was consistently valued, including nurse-led monitoring [77], weekly coaching calls [80], and live remote supervision [81], while the absence of therapist support was experienced as a barrier [78]. Gamification remained largely facilitative but was not uniformly positive, as competitive leaderboards discouraged lower-ranked users [82]. Technical and logistical friction continued to undermine adherence, with device charging and synchronization problems and simply forgetting to wear trackers reported repeatedly [83,84], whereas embedding the intervention in a familiar platform and delivering it remotely helped overcome access and geographic barriers [77,80].

### *User-Related Factors*

User-related factors associated with adherence to physical activity interventions spanned demographic, motivational, and practical dimensions. Statistical evidence revealed that Hispanic ethnicity was positively associated with adherence [20]. Descriptive evidence identified gender-based motivational differences, with boys engaging primarily for enjoyment and girls participating to combat boredom [33], while injury, illness, and pain were associated with lower adherence [23]. Qualitative evidence highlighted that opportunities for family participation [28] and users' perception of personal benefits [26] enhanced adherence, whereas lack of interest in the intervention [26,27], time constraints [23,26], and short-term motivation [28] acted as barriers. Practical challenges, including limited access to Wi-Fi or mobile data [28], insufficient device storage [32], and simply forgetting to engage with the intervention [31,32], were also reported as obstacles to sustained use.

The new physical activity studies extended the user-related picture, particularly for motivation, age, and clinical burden. Intrinsic motivation and perceived benefit again facilitated engagement, including satisfaction of competence and autonomy needs [85] and perceived health gains [78,80], whereas waning motivation and loss of interest drove disengagement [77,80,81]. The direction of age was not uniform, as older participants reported a higher intention to reuse a virtual exercise program than younger participants [86]. A new and distinct theme was illness burden in clinical samples, where fatigue, pain, comorbidities, hospitalizations, and fear of hypoglycemia limited participation [79-81]. Time constraints from school and competing demands, together with the poor fit of frequent device charging with busy routines, were again reported as obstacles [77,80,83].

## Obesity Management

### *Intervention-Related Factors*

Several intervention design features were associated with adherence to digital obesity management interventions. Statistical evidence indicated that communication with a health coach [34,35], social voting polls [35], parental participation [35,36], online food journaling [37], reinforcement messages [38], primary recruitment sites [39], and in-class program delivery versus homework-based delivery [40] were significantly associated with increased adherence. Adherence was also significantly higher during the first half compared to the second half of interventions [36,41]. Descriptive evidence supported higher adherence rates associated with health coach communication [39,42], specific program features (My workouts, My steps, and My goal functions) [43], and the initial intervention phase [37,39,44]. Conversely, statistical evidence indicated that self-esteem and stress management message components [38] and the AppAlone program compared to the AppCoach program [39] were associated with reduced adherence. Descriptive evidence revealed barriers including the ineffective "my motivation" function [43] and the selected self-monitoring tool [36].

Qualitative evidence highlighted several facilitating features: informative videos covering topics such as nutrition label reading, heart rate measurement, and screen time health effects [45], daily encouragement with real-time feedback, gamification elements, peer competition, and accessible health care professional communication [46]. Users also positively perceived daily reminders, craving management strategies, and weekly coach phone calls [39]. However, qualitative evidence also revealed barriers including insufficient individualization and support, limited exercise variety, and inadequate in-person resources and group activities [47], technical difficulties [40,48], excessive and repetitive messaging [43], overly demanding training requirements regarding frequency and duration [48], and educator concerns about lesson length and reading volume [40].

### *User-Related Factors*

User-related factors associated with adherence to obesity management interventions encompassed demographic, psychological, and behavioral dimensions. Statistical evidence demonstrated positive associations between adherence and successful BMI reduction in the initial 12 weeks [47], female gender [40], and younger age [40,44]. Additional significant predictors included intrinsic motivation, healthier parenting practices, and higher household income [44]. Conversely, statistical evidence indicated lower adherence associated with authoritative parenting style and depression [41,44]. Descriptive evidence revealed contrasting findings regarding socioeconomic status, with one study suggesting that lower household income was associated with better adherence [39]. Further descriptive evidence identified greater weight and shape concerns and white ethnicity [37], higher weight status, interfering activities, and scheduling conflicts such as being too busy, tired, or away from home [48] as barriers to adherence. Qualitative evidence revealed lost motivation as an additional barrier [48].

## Dietary Habits

### *Intervention-Related Factors*

Several intervention design features were associated with adherence to dietary habit interventions. Statistical evidence



XSL•FO
RenderX

indicated that text message frequency of 2 - 3 messages per week was positively associated with adherence [49]. Descriptive evidence demonstrated positive effects from online resources and interactive quizzes [49], while a declining participation pattern was observed across the study duration [50]. Qualitative evidence supported the effectiveness of music integration, goal-setting functionalities, practical tips and recipes, food diary capabilities, and reminder systems [51,52], feedback mechanisms, device comfort, and visual design elements [52], as well as video content quality, language appropriateness, and overall layout structure [53]. However, qualitative evidence also revealed barriers including limited scenario applicability in dining hall settings and insufficient video content [53], inability to pause videos and excessive session duration [54], restricted reward options with specific requests for more song choices, and constrained food logging capabilities that only allowed tracking of fruits and vegetables [52].

Three new dietary studies reinforced the importance of tailoring, gamification, and content quality while highlighting reward-design pitfalls. Culturally and developmentally appropriate tailoring facilitated sustained use, including a slide-bar feature for shared family meals [87] and age-relevant rather than adult-oriented messaging [88]. Gamification elements such as badges, team goals, and social sharing were again linked to engagement [87], and credible, concise content was valued, although unmet requests for more recipes and features tempered enthusiasm [88]. Reward structures could backfire, however, as unattainable high-value rewards and delayed or embarrassing reward distribution discouraged engagement [89]. Technical and access constraints, including an Android-only app and limited school computer time, further limited use [89].

### *User-Related Factors*

User-related factors associated with adherence to dietary interventions encompassed behavioral, psychological, and practical dimensions. Statistical evidence indicated positive associations between adherence and higher education levels, healthy snack consumption patterns, elevated self-efficacy scores for healthy eating, and perceived competence in app usage [50]. Descriptive evidence showed positive relationships with lower sensitivity to reward scores and maintenance of adherence among users experiencing positive emotional responses to the app [50]. Descriptive evidence also indicated that time constraints due to academic workload and general lack of available time hindered adherence [52].

The new dietary evidence on user characteristics was consistent with the previously established picture. Excessive or poorly understood notifications and overly long assessments were experienced as barriers that competed with adolescents' limited time and attention [88]. Practical constraints on when and where the intervention could be used, such as restrictions on phone use at school, again shaped engagement [87,89].

## Tobacco Use

### *Intervention-Related Factors*

Few intervention design features were identified in relation to adherence to nicotine use interventions. Descriptive evidence indicated that ease of use, engaging content, and the presence of a narrator enhanced adherence [55]. Statistical evidence demonstrated significant participant dropout on the designated "quit day" [56], and qualitative evidence identified technological implementation issues as barriers to adherence [57].

The updated search markedly strengthened the tobacco evidence, which had been sparse, adding 5 studies that clarified the roles of prompts, peer support, platform integration, and third-party technical dependencies. Multiple well-timed prompts significantly raised engagement among historically underengaged users [90], and integrating the intervention into clinical workflows or the electronic health record supported high overall engagement [90], as did opting participants in during an in-person visit rather than through later automated texts [91]. Peer-delivered support was a recurring facilitator, including counselor-facilitated group chats that raised perceived support [92] and peer-written messages that generated the highest engagement [93], although reactive rather than proactive peer coaching limited engagement [93]. Dependence on third-party platforms introduced new technical barriers, as mobile carriers automatically blocked a large share of outreach texts as spam [91] and changes to social media rules and handles drove attrition [92]. Low production quality and a single storyline also constrained engagement with a peer-video intervention [93].

### *User-Related Factors*

User-related factors associated with adherence to nicotine use interventions were primarily identified through descriptive evidence. Facilitating characteristics included higher academic performance, regular school attendance, English as the home language, having internet access at home, and nonsmoking status [57]. Additional facilitators included extended prequit duration and multiple previous quit attempts (2 or more) [56]. Conversely, being a current or former smoker [57] and geographical location, specifically residing in the southern United States [56], were associated with lower adherence.

The new tobacco studies also enriched the user-related picture, which had previously rested on a single study, and introduced several within-domain contradictions. Higher baseline risk, including greater vaping frequency, dependence, and internalizing-symptom severity, predicted nonresponse [94], and non-susceptible never-smokers engaged more with prompts [90]. Motivation mattered, as low confidence to quit predicted nonresponse [94] and a quitting-centered framing did not resonate with adolescents who did not identify as smokers [91]. Gender associations pointed in opposite directions across studies, with male gender predicting nonresponse in one program [94] but female gender associated with greater attrition in another [92]. A distinctively adolescent barrier was privacy, as participants were reluctant to text a clinician while at school or near a parent, and home delivery of nicotine replacement risked forcing disclosure to parents [91].

## Alcohol Use

### *Intervention-Related Factors*

Few intervention design features were identified in relation to adherence to alcohol consumption interventions. Descriptive evidence supported that message frequency, particularly 1 - 3



XSL•FO
RenderX

messages per week, was associated with higher adherence [58]. Conversely, descriptive evidence showed that home-based booster sessions, compared to school settings, were associated with lower adherence [59]. Qualitative evidence revealed technical barriers impacting adherence, including ineffective email delivery resulting in messages being directed to spam folders and general logistical difficulties [59].

Three newly identified alcohol studies reinforced the roles of usability, tailoring, and prompts while adding contradicting evidence on self-monitoring. A positive overall evaluation of an eHealth game predicted higher usability ratings, and its attractive layout and gamified avatars, rewards, and stories were well received [95], while tailored feedback supported engagement in a web-based program [96]. A peer-to-peer nudge strongly predicted program completion [97]. Counterbalancing the facilitative view of self-monitoring, one program found no dose-response between content exposure and outcomes [96]. Over-long or repetitive elements, including film lectures and baseline questionnaires, and a logout-and-break structure that discouraged reengagement, were reported as barriers [95,96].

### *User-Related Factors*

User-related factors associated with adherence to alcohol consumption interventions spanned demographic and behavioral dimensions. Statistical evidence demonstrated higher adherence among females, younger individuals, and those with higher educational levels [60], as well as participants of Spanish nationality [59]. Descriptive evidence indicated additional facilitating characteristics, including younger age [58,59], Dutch nationality [60], female gender, and higher maximum drinks per occasion [58]. Furthermore, descriptive evidence suggested that lower weekly alcohol consumption was linked to better adherence [59,60]. Conversely, descriptive evidence identified barriers including commencement of workplace placements during the intervention period, logistical challenges related to school commitments, and lower educational attainment [59].

The newly identified alcohol studies broadened the user-related determinants, especially around baseline risk, family factors, and socioeconomic status. Contrary to the prior pattern, adolescents reporting binge drinking rated the program more positively and engaged more with its action-planning content [95], while depression was linked to low adherence in a highly distressed sample [96]. Family and socioeconomic factors were prominent, as parental communication self-efficacy and open communication predicted completion and retention, higher household income raised initiation, and full-time parental employment aided retention [97]. Age and gender associations were mixed, with younger teens showing greater retention and completion [97] but older adolescents more likely to recommend the program [95]. Notably, self-identification with an ethnic category other than White was associated with greater retention [97], a facilitator direction that contrasts with the barrier role of minority status seen in the tobacco domain.

## Multicomponent Lifestyle Interventions

### *Intervention-Related Factors*

Several intervention design features were associated with adherence to digital multicomponent lifestyle interventions. Statistical evidence indicated that high parent engagement [61], self-regulation and narrative components [62], and reminder messages [63] were significantly associated with increased adherence. Descriptive evidence highlighted several facilitating factors: comfortable equipment [64], easy handling [65], incentives, access to new information, and progress feedback [66]. Additional positive effects were observed from point systems [67], quizzes [68,69], recipe provision [68], SMS prompts, prizes, and 4-month program duration [69,70]. The social skills component showed high participation rates [69,70], with peak interaction times observed between 3 PM and 1 AM on weekdays and weekends [71]. Conversely, statistical evidence demonstrated reduced adherence associated with smoking cessation stage-of-change assessments and contest prompts [72], as well as picture and message contests [69,70]. Descriptive evidence identified barriers including uncomfortable heart rate monitors [64], limited new website content [66], and declining participation rates over extended study durations [61,68,71].

Qualitative evidence supported the effectiveness of a concise baseline questionnaire and platform design [73], friendly health competitions and small altruistic rewards [74], and motivational support features [75]. However, qualitative evidence also revealed challenges related to poorly timed surveys [64], time-intensive baseline questionnaires, and technical difficulties encountered during the intervention [75].

Twelve newly identified multicomponent studies substantially expanded this domain and, in several cases, complicated the earlier facilitator picture. Gamification and reward mechanics were again linked to engagement, including health coins, streaks, and unlockable content associated with high weekly logins [98] and smartwatch reward displays that motivated use [99], whereas the absence of gamification was experienced as a barrier [106]. The clearest complication concerned self-monitoring: although tracking with feedback facilitated engagement in some programs [99,100], the perceived burden of daily self-monitoring reduced adherence and conferred no advantage over paper records in one trial [101], and self-monitoring components saw low or only partial uptake elsewhere [102,103,110]. Parental involvement remained facilitative, with caregiver-adolescent engagement correlated [104] and a mother a more acceptable co-participant than a friend [105]. Engagement typically declined over time, with effects lost by 18 months and most participants stopping after a few weeks [105,106], while shorter duration and completion incentives reduced dropout [107]. Incentives showed mixed effects, raising short-term motivation that faded on withdrawal [104] and, when coercively grade-linked, degrading app perception with heavier use [108].

### *User-Related Factors*

User-related factors associated with adherence to multicomponent lifestyle interventions were extensive. Statistical evidence showed higher adherence among females, individuals with weaker intentions to consume fruits and vegetables, and British nationals [73], as well as those with less exercise at baseline, lower BMI, and better health status [75]. Additional statistical predictors included mothers versus fathers in parent-child dyads and parents with high education [61], younger participants [76], those who had been drunk in the past



XSL•FO
RenderX

month, and previous users of the internet for health purposes [63], as well as individuals with higher perceived benefits of quitting [72]. Descriptive evidence indicated increased adherence among males [74], individuals with higher risk perception of drugs, lower alcohol and tobacco consumption, and fewer family-related alcohol and drug abuse problems [63], as well as younger participants, females, and those with nonproblematic drinking patterns [69,70]. Qualitative evidence highlighted enhanced adherence when participants felt part of a larger initiative and enjoyed learning about health and engaging in personal challenges [64].

Conversely, statistical evidence identified lower adherence among adolescents from nonacademic educational tracks compared to academic tracks [62], those with religious affiliation [63], and individuals with older age, binge drinking behavior, and immigration background [72]. Descriptive evidence revealed barriers including time constraints [66], perceived lack of behavioral change when using the app, subjective perception of already maintaining a healthy lifestyle, and insufficient environmental motivation [62], concurrent use of alternative lifestyle tracking apps and channels [62,63], lost login credentials [68], lack of interest [63,68], engagement in extracurricular activities, school absences, relationship status, and preference for alternative communication channels [63], and immigration background [69,70]. Qualitative evidence highlighted additional barriers including time limitations [74], insufficient device storage capacity, and the perception of mandatory participation [62].

The newly identified multicomponent studies also enriched the user-related determinants and reinforced several barriers. Lower socioeconomic position was again implicated, as noncompleters had significantly lower paternal education [101]. Loss of interest and waning motivation were among the most common reasons for dropout [98,101,106], and competing demands on adolescents' time, including school examinations, travel distance, and parental work, repeatedly limited engagement [99,101,103,110]. Access and device constraints, including device dependency and exclusion of adolescents without smartphones, raised digital-divide concerns [98,99]. A recurring adolescent-specific theme was privacy, which cut in both directions: a private, user-controlled interface reduced fear of judgment [101], yet a program that felt too private discouraged involving a friend and some adolescents preferred passive, nonsocial participation [102,105]. A late-responder subgroup with a mixed risk profile further illustrated that engagement timing varies across adolescents [109].

Twenty-six of the included studies contributed only to the descriptive synthesis of study characteristics and, where attrition data were extractable, to the quantitative synthesis of attrition; no factors influencing adherence were reported in these studies, and they therefore do not appear in the factor-level synthesis presented in Tables S1-S6 Multimedia Appendix 1. They comprised studies of physical activity [119-125], dietary habits [126], obesity management [127-130], tobacco use [131-136], alcohol use [137-140], multicomponent lifestyle interventions [141-143], and the study by Parker et al [147].

# Discussion

## Principal Findings

This systematic review synthesized evidence from 116 studies to identify the factors influencing adherence to digital lifestyle interventions for adolescents. Our analysis revealed that, although just over half (47/89, 52.8%) of the interventions successfully achieved their primary behavioral outcomes, this effectiveness is likely to be undermined by significant adherence challenges. On average, participants completed 66.9% (SD 26%) of the interventions (the mean completion rate across the 68 comparisons that reported it). Mean adherence to individual intervention components was lower still: 55.2% (SD 25.5%) (across the 66 comparisons that reported it). This disparity highlights a critical issue, namely that an intervention is unlikely to be effective if users do not engage with it as intended [148]. Therefore, understanding the drivers of adherence is crucial.

## Intervention-Related Determinants of Adherence

Several intervention-related elements were consistently effective in promoting adherence across multiple health domains. Personalization capabilities, intuitive user interfaces, strategically timed reminder systems, goal-setting functionalities, and gamification elements were found to be particularly influential design features. These findings align with and extend research on adult populations. For instance, the importance of personalization and user-friendly design echoes findings from systematic reviews on digital mental health interventions for adults [10]. However, our review indicates that for adolescents, these elements may need to be combined with gamification and engaging activities to be successful. This is consistent with self-determination theory, which posits that supporting autonomy (personalization), competence (usability and goal-setting), and relatedness (human support) fosters the intrinsic motivation crucial for sustained engagement [5,149]. The effectiveness of gamification elements, such as music integration, interactive quizzes, and competitive features, suggests that interventions that leverage intrinsic motivation through enjoyable activities may achieve better adherence outcomes than purely informational approaches [150]. Importantly, while gamification and "engaging activities" are related, they are operationally distinct in the included studies. Gamification refers to the application of game-design elements (eg, points, badges, leaderboards, and competition) to nongame contexts [151], whereas engaging activities encompass a broader category of design features intended to sustain interest, including interactive quizzes, music integration, narrative storylines, and varied content formats. Both contributed to adherence, but through different mechanisms: gamification primarily leverages extrinsic motivational pathways through reward and competition, while engaging activities may support intrinsic engagement through enjoyment and curiosity.

The prominent role of human support elements, which is particularly evident in obesity interventions through health coach communication and parental involvement, highlights the ongoing importance of interpersonal connections in digital health contexts [5,152]. This finding supports the emphasis placed on social support as a critical determinant of behavior



XSL•FO
RenderX

change maintenance in social cognitive theory [153,154]. However, the mixed evidence regarding the effectiveness of social intervention components, such as message boards and expert forums, suggests that not all forms of social interaction enhance adherence. The effectiveness of human support may depend on the quality, timing, and perceived relevance of the interaction, rather than on the mere availability of social features.

Implementing comprehensive human support presents significant scalability challenges, particularly for universal prevention programs, where providing individual attention becomes logistically and economically unfeasible. Constraints on resources in health care and education systems often prevent the provision of sustained professional support for large groups of participants. While hybrid models combining digital tools with personal support can optimize adherence in clinical or targeted prevention contexts, alternative strategies deserve consideration for population-level implementations. These could include peer mentoring systems, AI-driven personalized feedback mechanisms, or tiered support models where human interaction is strategically deployed for participants showing signs of early disengagement or elevated risk [155].

Technical implementation barriers consistently emerged as substantial obstacles to sustained adherence in all health domains, regardless of the focus of the specific intervention. Synchronization difficulties, authentication problems, and inappropriately timed notifications emphasize the importance of robust technological infrastructure and comprehensive, user-centered design principles. These technical challenges were particularly evident in interventions targeting physical activity and alcohol use, potentially due to the real-time tracking requirements and environmental context sensitivity inherent to these health domains. The consistent finding of declining adherence over the duration of interventions [156], particularly in dietary and multicomponent domains, suggests the necessity of adaptive engagement strategies that evolve dynamically to maintain user interest and combat habituation effects.

### User-Related Determinants of Adherence

The synthesis revealed complex and occasionally contradictory relationships between demographic characteristics and adherence patterns. This indicates that associations are context-dependent and cannot be easily generalized. While some studies identified higher adherence among female participants and younger adolescents, these relationships demonstrated considerable variability across health domains and study contexts. This suggests that demographic predictors may be moderated by intervention characteristics, the health domain in question, and cultural factors. This heterogeneity is consistent with ecological systems theory, which emphasizes the multifaceted influences on adolescent behavior, including individual, family, peer, and broader societal factors [157].

Socioeconomic indicators, particularly educational attainment and household income, exhibited more consistent associations with adherence; however, the nature of these relationships varied across health domains. Higher educational levels were associated with better adherence to dietary and alcohol interventions, which may reflect enhanced health literacy and digital competency. However, conflicting evidence regarding the effects of household income on obesity interventions was found, suggesting that socioeconomic influences may operate differently across behavioral targets. These findings emphasize the importance of considering the implications for health equity in the design of digital interventions and the potential need for differentiated implementation strategies across socioeconomic groups.

These socioeconomic patterns in adherence have important implications for health equity that extend beyond the adolescent-specific literature examined here. Systematic reviews examining digital health interventions in adult populations have identified similar socioeconomic gradients in engagement. Szinay et al [158] found that the uptake, engagement with, and effectiveness of mobile interventions for weight-related behaviors were not equally distributed across socioeconomic groups, raising concerns that digital health interventions may inadvertently widen health inequalities. Western et al [159] further argued that the “digital health divide” operates at multiple levels, including individual digital literacy, infrastructure access, and the socioeconomic context of health domains. Our findings in adolescent populations echo these adult-focused reviews and suggest that the equity implications of digital health interventions may emerge early in the lifespan. Interventions targeting adolescents from lower socioeconomic backgrounds may require differentiated design and implementation strategies, such as offline functionality, reduced data requirements, or integration with existing community infrastructure, to ensure equitable engagement.

Psychological factors, particularly intrinsic motivation and the perceived personal benefits of the intervention, emerged as powerful and consistent facilitators. This is particularly relevant for adolescents, who, as noted in the literature, are in a critical period of identity formation during which their cognitive decision-making and critical thinking functions are still maturing [11,12]. Interventions that do not align with their developing personal values or are not perceived as immediately relevant or beneficial are likely to be abandoned, which is consistent with motivation theories that emphasize the need for perceived efficacy and value alignment to encourage long-term participation [149,160]. However, the frequent reporting of loss of motivation and declining interest as barriers to adherence, particularly in the obesity and physical activity domains, highlights the difficulty of maintaining initial enthusiasm throughout extended intervention periods.

### Domain-Specific Adherence Patterns

Adherence-related outcomes varied across health domains, illuminating the different challenges and motivators associated with specific behavioral targets. In the quantitative synthesis of attrition, dropout was lowest in physical activity interventions, potentially reflecting the immediate feedback and tangible outcomes associated with movement tracking, and highest in alcohol and tobacco interventions, possibly reflecting the sensitive nature of substance-use disclosure and the complex psychological factors surrounding these behaviors in adolescents. Completion and component-adherence rates also differed across domains, although the wide and overlapping



XSL•FO
RenderX

ranges caution against overinterpreting any single domain ordering.

Multicomponent lifestyle interventions illustrate the tension between comprehensiveness and adherence. While targeting multiple risk factors simultaneously is theoretically appealing, the expanded evidence base indicated that such programs frequently encountered engagement decline, self-monitoring burden, and competing demands, suggesting that they may present unique adherence challenges that require specialized design considerations to maintain participant involvement across diverse behavioral targets.

Quantitatively synthesizing attrition reinforced and sharpened the narrative findings. Although a substantial minority of adolescents (pooled 16.9%) dropped out of the included studies, the between-study heterogeneity was extreme ($I^2$>98%) and was not explained by health domain, delivery modality, human support, incentives, intervention duration, or participant age. The level of disengagement is highly context-specific and is not predicted by the broad design characteristics that are typically reported, which supports our central recommendation that the field should adopt standardized adherence metrics. The contrast between modest attrition (approximately 17%) and lower intervention use, with mean completion of approximately two-thirds and mean component adherence of approximately one-half, is equally instructive. Adolescents tend to remain nominally enrolled while ceasing to engage with the intervention as intended, so that retention statistics alone substantially overstate true adherence.

These quantitative patterns are consistent with recent syntheses reporting low and declining real-world engagement with digital interventions among young people and a comparable facilitator and barrier structure [161-164].

## Limitations

This systematic review has several important limitations. First, the study protocol was not prospectively registered, and we did not conduct forward or backward citation searching, which may have resulted in the omission of relevant studies not indexed in the searched databases. Second, we did not conduct formal risk of bias assessments of included studies. While validated tools exist for assessing risk of bias across different study designs (eg, the revised Cochrane Risk of Bias tool for randomized trials [RoB 2] [165] and the Risk Of Bias In Non-randomized Studies of Interventions [ROBINS-I] tool [166]), our primary objective was to identify and categorize factors reported as influencing adherence rather than to evaluate intervention effectiveness. Because we do not aggregate effect estimates or draw causal conclusions about intervention efficacy, a formal risk-of-bias assessment of each study's primary outcomes would not directly inform our narrative synthesis. We acknowledge, however, that this means findings from studies with varying methodological rigor receive equal weight in our synthesis. Third, this review focused on 5 lifestyle-related health domains selected based on their relevance to cancer prevention. The findings may not be generalizable to digital interventions targeting other health domains, such as mental health or sexual health, where the developmental and behavioral contexts differ substantially. Finally, although attrition could be meta-analyzed, the pooled estimates should be interpreted in light of very high between-study heterogeneity and the inconsistent operationalization of attrition and adherence across studies; the estimates are descriptive, are not weighted by study-level risk of bias, and the small-study-effect test should be read cautiously at this level of heterogeneity. Completion and component-level adherence were reported too heterogeneously to support a defensible pooled estimate and were therefore synthesized descriptively and through the SWiM approach rather than meta-analytically.

## Implications for Future Research

The findings of this review have several implications for future research and practice. First, the considerable heterogeneity in adherence measurement and reporting represents a fundamental barrier to advancing the field. We recommend that future studies adopt standardized adherence metrics, distinguishing at minimum between completion rates (proportion completing the intervention) and component-level adherence (engagement with specific intervention elements), and report both measures alongside their operationalization. Second, the consistent role of intervention design features across multiple health domains suggests that these elements should be considered core components in the design of digital health interventions for adolescents. Third, the mixed evidence regarding social features highlights the need for experimental research that systematically varies the type, timing, and intensity of social components to determine optimal configurations for different target groups. Fourth, the clear impact of technical barriers across all domains underscores the importance of rigorous usability testing and iterative design processes in intervention development. Finally, future research should explicitly examine the dose-response relationship between adherence and intervention effectiveness. Meta-analyses of individual participant data could provide enhanced statistical power for identifying optimal intervention parameters while accounting for participant-level moderators of adherence [167].

## Conclusion

This systematic review provides a comprehensive overview of the factors influencing adherence to digital lifestyle interventions among adolescents. The findings show that intervention-related design features and user-related characteristics consistently affect adherence patterns, although these relationships are often context- and domain-specific. Complementing this narrative synthesis, a random-effects meta-analysis of attrition across 108 comparisons produced a pooled dropout of approximately 17%, but with extreme between-study heterogeneity that none of the examined study-level moderators explained, reinforcing that the degree of disengagement is highly context-specific rather than predictable from broad design features. The substantial variation in adherence rates and measurement approaches highlights the urgent need to establish standardized reporting frameworks that incorporate validated measures and clearly defined usage parameters. Such methodological advancements are essential for synthesizing robust evidence and translating research findings into effective, engaging digital health interventions for adolescent populations.



XSL•FO
RenderX

## Acknowledgments

The authors would like to acknowledge the researchers who conducted the primary studies and systematic reviews that form the evidence base for this work. We also extend our gratitude to the adolescent participants whose involvement in the original studies made this synthesis possible.

No generative AI tools were used to design the study, conduct the analyses, interpret the findings, generate scientific content, or identify references. Because some authors are not native English speakers, DeepL was used to assist with translation and grammatical and stylistic language editing. All resulting text was critically reviewed and revised by the authors, who take full responsibility for the accuracy and integrity of the manuscript.

## Funding

This study was supported by the SUNRISE (Sustainable Interventions and Healthy Behaviours for Adolescent Primary Prevention of Cancer with Digital Tools) project that has received funding from the European Union's Horizon Europe research and innovation program under Grant Agreement Number 101136829. Furthermore, this work received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Authors' Contributions

Conceptualization: NB (lead), SH (lead)
Methodology: NB (lead), SH (lead)
Formal analysis: NB (lead)
Investigation: NB, OS, PdR, IDY
Validation: NB, OS, PdR, IDY
Project administration: NB, SH
Supervision: SH (lead), NB, MPS (lead)
Funding acquisition: AT, SH
Visualization: OS
Writing – original draft: NB (lead), OS (supporting)
Writing – review & editing: KK, KC, AMB, TdPP, HK, AT, SH, MPS
All authors have read and agreed to the published version of the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies, factors influencing adherence by health domain (Tables S1-S6), characteristics and adherence-related characteristics of the included studies (Tables S7-S18), and the Synthesis Without Meta-analysis effect-direction table.
[DOCX File, 701 KB - i-jmr_v15i1e84822_app1.docx ]

Checklist 1
PRISMA 2020 checklist.
[PDF File, 144 KB - i-jmr_v15i1e84822_app2.pdf ]

## References

1. de Sousa D, Fogel A, Azevedo J, Padrão P. The effectiveness of web-based interventions to promote health behaviour change in adolescents: a systematic review. Nutrients 2022 Mar 16;14(6):1258. [doi: 10.3390/nu14061258] [Medline: 35334915]
2. Rose T, Barker M, Maria Jacob C, et al. A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. J Adolesc Health 2017 Dec;61(6):669-677. [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
3. Champion KE, Parmenter B, McGowan C, et al. Effectiveness of school-based eHealth interventions to prevent multiple lifestyle risk behaviours among adolescents: a systematic review and meta-analysis. Lancet Digit Health 2019 Sep;1(5):e206-e221. [doi: 10.1016/S2589-7500(19)30088-3] [Medline: 33323269]
4. Partridge SR, Redfern J. Strategies to engage adolescents in digital health interventions for obesity prevention and management. Health Care (Basel) 2018;6(3):70. [doi: 10.3390/healthcare6030070]



XSL•FO
RenderX

5. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
6. Kernebeck S, Busse TS, Ehlers JP, Vollmar HC. Adherence to digital health interventions: definitions, methods, and open questions. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021 Oct;64(10):1278-1284. [doi: 10.1007/s00103-021-03415-9] [Medline: 34559252]
7. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
8. Short CE, DeSmet A, Woods C, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292. [doi: 10.2196/jmir.9397] [Medline: 30446482]
9. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
10. Hornstein S, Zantvoort K, Lueken U, Funk B, Hilbert K. Personalization strategies in digital mental health interventions: a systematic review and conceptual framework for depressive symptoms. Front Digit Health 2023;5:1170002. [doi: 10.3389/fdgth.2023.1170002] [Medline: 37283721]
11. Zimmermann E, Tomczyk S. Using social media to promote life skills among adolescents: a debate on opportunities, challenges, and implications for health and education. J Prev 2025 Apr;46(2):201-211. [doi: 10.1007/s10935-025-00826-1]
12. Shanmugasundaram M, Tamilarasu A. The impact of digital technology, social media, and artificial intelligence on cognitive functions: a review. Front Cognit 2023;2. [doi: 10.3389/fcogn.2023.1203077]
13. Nyman J, Tornivuori A, Salanterä S, Barroso T, Parisod H. Systematic review of digital interventions to support refusal self-efficacy in child and adolescent health promotion. Health Promot Int 2022 Oct 1;37(5):daac085. [doi: 10.1093/heapro/daac085] [Medline: 36166268]
14. Lou J, Wang M, Xie X, et al. The association between family socio-demographic factors, parental mediation and adolescents' digital literacy: a cross-sectional study. BMC Public Health 2024 Oct 23;24(1):2932. [doi: 10.1186/s12889-024-20284-4] [Medline: 39438851]
15. Liu C, Messer M, Linardon J, Fuller-Tyszkiewicz M. Applying models of self-regulated learning to understand engagement with digital health interventions: a narrative review. Front Digit Health 2025;7:1380088. [doi: 10.3389/fdgth.2025.1380088] [Medline: 40416095]
16. El-Nasr MS, Durga S, Shiyko M, Sceppa C. Unpacking adherence and engagement in pervasive health games. arXiv. Preprint posted online on Jun 25, 2021. [doi: 10.48550/arXiv.2106.13747]
17. Oh C, Carducci B, Vaivada T, Bhutta ZA. Digital interventions for universal health promotion in children and adolescents: a systematic review. Pediatrics 2022 May 1;149(Suppl 5):e2021053852H. [doi: 10.1542/peds.2021-053852H] [Medline: 35503335]
18. Yang X, Ma L, Zhao X, Kankanhalli A. Factors influencing user's adherence to physical activity applications: a scoping literature review and future directions. Int J Med Inform 2020 Feb;134:104039. [doi: 10.1016/j.ijmedinf.2019.104039] [Medline: 31865054]
19. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
20. Dunton GF, Dzubur E, Intille S. Feasibility and performance test of a real-time sensor-informed context-sensitive ecological momentary assessment to capture physical activity. J Med Internet Res 2016 Jun 1;18(6):e106. [doi: 10.2196/jmir.5398] [Medline: 27251313]
21. Tripicchio GL, Jones GJ, Hart CN, et al. A digitally enhanced home-based physical activity intervention for high-risk middle school youth during COVID-19. Transl Behav Med 2023 Jan 20;13(1):17-24. [doi: 10.1093/tbm/ibab151] [Medline: 34850218]
22. Kaos MD, Beauchamp MR, Bursick S, et al. Efficacy of online multi-player versus single-player exergames on adherence behaviors among children: a nonrandomized control trial. Ann Behav Med 2018 Sep 13;52(10):878-889. [doi: 10.1093/abm/kax061] [Medline: 30212848]
23. Heale LD, Dover S, Goh YI, Maksymiuk VA, Wells GD, Feldman BM. A wearable activity tracker intervention for promoting physical activity in adolescents with juvenile idiopathic arthritis: a pilot study. Pediatr Rheumatol Online J 2018 Oct 22;16(1):66. [doi: 10.1186/s12969-018-0282-5] [Medline: 30348203]
24. Knox E, Glazebrook C, Randell T, et al. SKIP (Supporting Kids with diabetes In Physical activity): feasibility of a randomised controlled trial of a digital intervention for 9-12 year olds with type 1 diabetes mellitus. BMC Public Health 2019 Apr 3;19(1):371. [doi: 10.1186/s12889-019-6697-1] [Medline: 30943939]
25. Stephens S, Schneiderman JE, Finlayson M, Berenbaum T, Motl RW, Yeh EA. Feasibility of a theory-informed mobile app for changing physical activity in youth with multiple sclerosis. Mult Scler Relat Disord 2022 Feb;58:103467. [doi: 10.1016/j.msard.2021.103467] [Medline: 34954651]




XSL•FO
RenderX

26. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for IMproving FITness and increasing physical activity among young people: the AIMFIT pragmatic randomized controlled trial. J Med Internet Res 2015 Aug 27;17(8):e210. [doi: 10.2196/jmir.4568] [Medline: 26316499]
27. Ha L, Wakefield CE, Mizrahi D, et al. A digital educational intervention with wearable activity trackers to support health behaviors among childhood cancer survivors: pilot feasibility and acceptability study. JMIR Cancer 2022 Aug 17;8(3):e38367. [doi: 10.2196/38367] [Medline: 35976683]
28. Koorts H, Salmon J, Timperio A, et al. Translatability of a wearable technology intervention to increase adolescent physical activity: mixed methods implementation evaluation. J Med Internet Res 2020 Aug 7;22(8):e13573. [doi: 10.2196/13573] [Medline: 32763872]
29. Larsen B, Benitez T, Cano M, et al. Web-based physical activity intervention for Latina adolescents: feasibility, acceptability, and potential efficacy of the Niñas Saludables study. J Med Internet Res 2018 May 9;20(5):e170. [doi: 10.2196/jmir.9206] [Medline: 29743151]
30. Cushing CC, Bejarano CM, Ortega A, Sayre N, Fedele DA, Smyth JM. Adaptive mHealth intervention for adolescent physical activity promotion. J Pediatr Psychol 2021 Jun 3;46(5):536-546. [doi: 10.1093/jpepsy/jsaa125] [Medline: 33484137]
31. Pope L, Garnett B, Dibble M. Lessons learned through the implementation of an eHealth physical activity gaming intervention with high school youth. Games Health J 2018 Apr;7(2):136-142. [doi: 10.1089/g4h.2017.0164] [Medline: 29393679]
32. Zapata KA, Wang-Price SS, Fletcher TS, Johnston CE. Factors influencing adherence to an app-based exercise program in adolescents with painful hyperkyphosis. Scoliosis Spinal Disord 2018;13(1):11. [doi: 10.1186/s13013-018-0159-x] [Medline: 30027121]
33. Martínez-López EJ, López-Serrano S, De La Torre-Cruz M, Ruiz-Ariza A. Effects of the augmented reality game Pokémon GO on fitness and fatness in secondary school students. Health Educ J 2022 Feb;81(1):54-68. [doi: 10.1177/00178969211047800]
34. Bent MA, Valenzuela-Moss JN, Smith KA, et al. Using a mobile health app in a weight management program for overweight and obese adolescents and young adults with spina bifida. J Spinal Cord Med 2024 Nov;47(6):977-986. [doi: 10.1080/10790268.2023.2231675] [Medline: 37769142]
35. Deslippe AL, González ODJ, Buckler EJ, et al. Do individual characteristics and social support increase children's use of an mHealth intervention? Findings from the evaluation of a behavior change mHealth app, Aim2Be. Child Obes 2023 Oct;19(7):435-442. [doi: 10.1089/chi.2022.0055] [Medline: 36576875]
36. Tu AW, Watts AW, Chanoine JP, et al. Does parental and adolescent participation in an e-health lifestyle modification intervention improves weight outcomes? BMC Public Health 2017 Apr 24;17(1):352. [doi: 10.1186/s12889-017-4220-0] [Medline: 28438202]
37. Jones M, Luce KH, Osborne MI, et al. Randomized, controlled trial of an internet-facilitated intervention for reducing binge eating and overweight in adolescents. Pediatrics 2008 Mar;121(3):453-462. [doi: 10.1542/peds.2007-1173] [Medline: 18310192]
38. Kornman KP, Shrewsbury VA, Chou AC, et al. Electronic therapeutic contact for adolescent weight management: the Loozit study. Telemed E-Health 2010 Jul;16(6):678-685. [doi: 10.1089/tmj.2009.0180]
39. Vidmar AP, Salvy SJ, Wee CP, et al. An addiction-based digital weight loss intervention: a multi-centre randomized controlled trial. Pediatr Obes 2023 Mar;18(3):3. [doi: 10.1111/ijpo.12990] [Medline: 36484235]
40. Whittemore R, Chao A, Jang M, et al. Implementation of a school-based internet obesity prevention program for adolescents. J Nutr Educ Behav 2013;45(6):586-594. [doi: 10.1016/j.jneb.2013.03.012] [Medline: 23850014]
41. White MA, Martin PD, Newton RL, et al. Mediators of weight loss in a family‐based intervention presented over the internet. Obes Res 2004 Jul;12(7):1050-1059. [doi: 10.1038/oby.2004.132] [Medline: 15292468]
42. Chew CSE, Davis C, Lim JKE, et al. Use of a mobile lifestyle intervention app as an early intervention for adolescents with obesity: single-cohort study. J Med Internet Res 2021 Sep 28;23(9):e20520. [doi: 10.2196/20520] [Medline: 34581672]
43. Lubans DR, Smith JJ, Skinner G, Morgan PJ. Development and implementation of a smartphone application to promote physical activity and reduce screen-time in adolescent boys. Front Public Health 2014;2:42. [doi: 10.3389/fpubh.2014.00042] [Medline: 24904909]
44. Mâsse LC, Watts AW, Barr SI, et al. Individual and household predictors of adolescents' adherence to a web-based intervention. Ann Behav Med 2015 Jun;49(3):371-383. [doi: 10.1007/s12160-014-9658-z] [Medline: 25270826]
45. Furlong MC. Student education addressing pediatric obesity in maui county middle schools. ScholarSpace. 2021. URL: http://hdl.handle.net/10125/76457 [accessed 2025-01-29]
46. Stasinaki A, Büchter D, Shih CHI, et al. Effects of a novel mobile health intervention compared to a multi-component behaviour changing program on body mass index, physical capacities and stress parameters in adolescents with obesity: a randomized controlled trial. BMC Pediatr 2021 Jul 9;21(1):308. [doi: 10.1186/s12887-021-02781-2] [Medline: 34243738]
47. Bauer S, de Niet J, Timman R, Kordy H. Enhancement of care through self-monitoring and tailored feedback via text messaging and their use in the treatment of childhood overweight. Patient Educ Couns 2010 Jun;79(3):315-319. [doi: 10.1016/j.pec.2010.03.014] [Medline: 20418046]




XSL•FO
RenderX

48. Naets T, Vervoort L, Tanghe A, Braet C. Adherence and barriers in e-health self-control training for enhancing childhood multidisciplinary obesity treatment. Clin Psychol Psychother 2020 Jan;27(1):42-51. [doi: 10.1002/cpp.2405] [Medline: 31711275]
49. Haas K, Martin A, Park KT. Text Message Intervention (TEACH) improves quality of life and patient activation in celiac disease: a randomized clinical trial. J Pediatr 2017 Jun;185:62-67. [doi: 10.1016/j.jpeds.2017.02.062] [Medline: 28343658]
50. De Cock N, Van Lippevelde W, Vangeel J, et al. Feasibility and impact study of a reward-based mobile application to improve adolescents' snacking habits. Public Health Nutr 2018 Aug;21(12):2329-2344. [doi: 10.1017/S1368980018000678] [Medline: 29607800]
51. Hutcheson TD. Using mobile technology to impact fruit and vegetable consumption in low-income youth [dissertation]. : ProQuest Information & Learning; 2013 URL: https://kuscholarworks.ku.edu/entities/publication/3e711d7a-148f-4fc4-89ef-d41f2e6e7d84 [accessed 2026-09-10]
52. Nollen NL, Hutcheson T, Carlson S, et al. Development and functionality of a handheld computer program to improve fruit and vegetable intake among low-income youth. Health Educ Res 2013 Apr;28(2):249-264. [doi: 10.1093/her/cys099] [Medline: 22949499]
53. Monroe JT, Lofgren IE, Sartini BL, Greene GW. The Green Eating Project: web-based intervention to promote environmentally conscious eating behaviours in US university students. Public Health Nutr 2015 Sep;18(13):2368-2378. [doi: 10.1017/S1368980015002396] [Medline: 26282525]
54. Nicolaou P, Merwin RM, Karekla M. Acceptability and feasibility of a gamified digital eating disorder early-intervention program (AcceptME) based on Acceptance and Commitment Therapy (ACT). J Contextual Behav Sci 2022 Jul;25:26-34. [doi: 10.1016/j.jcbs.2022.06.001]
55. Braciszewski JM, Colby SM, Franklin MJ, Stout RL, Vose-O'Neal A, Bock BC. Technology-based smoking cessation for youth exiting foster care: a pilot randomized trial. J Public Child Welf 2024;18(1):61-79. [doi: 10.1080/15548732.2022.2155754] [Medline: 38144460]
56. Chan C, Kamke K, Assuah F, El-Toukhy S. Dropout, response, and abstinence outcomes of a national text-messaging smoking cessation intervention for teens, SmokeFreeTeen. Transl Behav Med 2021 Apr 7;11(3):764-771. [doi: 10.1093/tbm/ibaa069] [Medline: 32710628]
57. Buller DB, Borland R, Woodall WG, et al. Randomized trials on consider this, a tailored, internet-delivered smoking prevention program for adolescents. Health Educ Behav 2008 Apr;35(2):260-281. [doi: 10.1177/1090198106288982] [Medline: 17114331]
58. Haug S, Paz Castro R, Kowatsch T, Filler A, Dey M, Schaub MP. Efficacy of a web- and text messaging-based intervention to reduce problem drinking in adolescents: results of a cluster-randomized controlled trial. J Consult Clin Psychol 2017 Feb;85(2):147-159. [doi: 10.1037/ccp0000138] [Medline: 27606700]
59. Martinez-Montilla JM, Mercken L, de Vries H, Candel M, Lima-Rodríguez JS, Lima-Serrano M. A web-based, computer-tailored intervention to reduce alcohol consumption and binge drinking among Spanish adolescents: cluster randomized controlled trial. J Med Internet Res 2020 Jan 24;22(1):e15438. [doi: 10.2196/15438] [Medline: 32012064]
60. Jander A, Crutzen R, Mercken L, Candel M, de Vries H. Effects of a web-based computer-tailored game to reduce binge drinking among Dutch adolescents: a cluster randomized controlled trial. J Med Internet Res 2016 Feb 3;18(2):e29. [doi: 10.2196/jmir.4708] [Medline: 26842694]
61. Lin Y, Mâsse LC. A look at engagement profiles and behavior change: a profile analysis examining engagement with the Aim2Be lifestyle behavior modification app for teens and their families. Prev Med Rep 2021 Dec;24:101565. [doi: 10.1016/j.pmedr.2021.101565] [Medline: 34976631]
62. Maenhout L, Peuters C, Cardon G, Crombez G, DeSmet A, Compernolle S. Nonusage attrition of adolescents in an mHealth promotion intervention and the role of socioeconomic status: secondary analysis of a 2-arm cluster-controlled trial. JMIR Mhealth Uhealth 2022 May 10;10(5):e36404. [doi: 10.2196/36404] [Medline: 35536640]
63. Balsa AI, Gandelman N, Lamé D. Lessons from participation in a web-based substance use preventive program in Uruguay. J Child Adolesc Subst Abuse 2014 Mar 1;23(2):91-100. [doi: 10.1080/1067828X.2012.748600]
64. Brannon EE, Cushing CC, Crick CJ, Mitchell TB. The promise of wearable sensors and ecological momentary assessment measures for dynamical systems modeling in adolescents: a feasibility and acceptability study. Transl Behav Med 2016 Dec;6(4):558-565. [doi: 10.1007/s13142-016-0442-4] [Medline: 27678501]
65. Fuemmeler BF, Holzwarth E, Sheng Y, et al. Mila Blooms: a mobile phone application and behavioral intervention for promoting physical activity and a healthy diet among adolescent survivors of childhood cancer. Games Health J 2020 Aug;9(4):279-289. [doi: 10.1089/g4h.2019.0060] [Medline: 32391734]
66. DeBar LL, Dickerson J, Clarke G, Stevens VJ, Ritenbaugh C, Aickin M. Using a website to build community and enhance outcomes in a group, multi-component intervention promoting healthy diet and exercise in adolescents. J Pediatr Psychol 2009 Jun;34(5):539-550. [doi: 10.1093/jpepsy/jsn126] [Medline: 19091807]
67. Kramer EN, Chard CA, Walters K, Barr-Anderson DJ. Weight-dependent disparities in adolescent girls: the impact of a brief pilot intervention on exercise and healthy eater identity. Int J Environ Res Public Health 2018 Jul 4;15(7):1411. [doi: 10.3390/ijerph15071411] [Medline: 29973565]




XSL•FO
RenderX

68. Rydell SA, French SA, Fulkerson JA, et al. Use of a Web-based component of a nutrition and physical activity behavioral intervention with Girl Scouts. J Am Diet Assoc 2005 Sep;105(9):1447-1450. [doi: 10.1016/j.jada.2005.06.027] [Medline: 16129089]
69. Haug S, Paz Castro R, Wenger A, Schaub MP. A mobile phone-based life-skills training program for substance use prevention among adolescents: cluster-randomized controlled trial. JMIR Mhealth Uhealth 2021 Jul 13;9(7):e26951. [doi: 10.2196/26951] [Medline: 34255703]
70. Paz Castro R, Haug S, Wenger A, Schaub MP. Longer-term efficacy of a digital life-skills training for substance use prevention. Am J Prev Med 2022 Dec;63(6):944-953. [doi: 10.1016/j.amepre.2022.06.017] [Medline: 35985899]
71. Schwinn TM, Schinke SP, Hopkins J, Keller B, Liu X. An online drug abuse prevention program for adolescent girls: posttest and 1-year outcomes. J Youth Adolesc 2018 Mar;47(3):490-500. [doi: 10.1007/s10964-017-0714-4] [Medline: 28755247]
72. Paz Castro R, Haug S, Filler A, Kowatsch T, Schaub MP. Engagement within a mobile phone–based smoking cessation intervention for adolescents and its association with participant characteristics and outcomes. J Med Internet Res 2017 Nov 1;19(11):e356. [doi: 10.2196/jmir.7928] [Medline: 29092811]
73. Cameron D, Epton T, Norman P, et al. A theory-based online health behaviour intervention for new university students (U@Uni:LifeGuide): results from a repeat randomized controlled trial. Trials 2015 Dec 7;16:555. [doi: 10.1186/s13063-015-1092-4] [Medline: 26643917]
74. Egilsson E, Bjarnason R, Njardvik U. Usage and daily attrition of a smartphone-based health behavior intervention: randomized controlled trial. JMIR Mhealth Uhealth 2023 Jun 26;11:e45414. [doi: 10.2196/45414] [Medline: 37358888]
75. Epton T, Norman P, Dadzie AS, et al. A theory-based online health behaviour intervention for new university students (U@Uni): results from a randomised controlled trial. BMC Public Health 2014 Jun 5;14(1):563. [doi: 10.1186/1471-2458-14-563] [Medline: 24903620]
76. Spook J, Paulussen T, Kok G, van Empelen P. Evaluation of a serious self-regulation game intervention for overweight-related behaviors ("Balance It"): a pilot study. J Med Internet Res 2016 Sep 26;18(9):e225. [doi: 10.2196/jmir.4964] [Medline: 27670222]
77. Li Y, Zhou Y, Zhang Y, et al. Effectiveness of a nurse-led, mHealth-based cardiac rehabilitation program on cardiorespiratory fitness and physical activity in youth with congenital heart disease: a randomized controlled trial. Int J Nurs Stud 2026 May;177:105370. [doi: 10.1016/j.ijnurstu.2026.105370] [Medline: 41690191]
78. Smith MH, Longmuir PE, Robb M, et al. Enhancing youth mental health through virtual lifestyle behavior change support: a pilot feasibility trial. Children (Basel) 2026 Jan 23;13(2):163. [doi: 10.3390/children13020163] [Medline: 41749519]
79. Hormazábal-Aguayo I, Muñoz-Pardeza J, Huerta-Uribe N, Ezzatvar Y, García-Hermoso A. Evaluating the usability of Diactive-1: mHealth for personalized exercise and education in children and adolescents with type 1 diabetes. Mhealth 2025;11:16. [doi: 10.21037/mhealth-24-63] [Medline: 40248752]
80. Miskin SA, V H, Naaz N, Alva R. m-Health-based exercise program among children and adolescents with chronic kidney disease: a feasibility study. Pediatr Nephrol 2026 Jun;41(6):1811-1820. [doi: 10.1007/s00467-026-07155-3] [Medline: 41580572]
81. Ozyemisci Taskiran O, Albayrak H, Kog C, et al. The effect of a 12-week tele-exercise using immersive virtual reality on functional capacity in adolescents with cystic fibrosis: a randomized controlled, single (assessor) - blind study. Respir Med 2025 Nov;248:108362. [doi: 10.1016/j.rmed.2025.108362] [Medline: 40983137]
82. Kwon S, Bai Y, Kim Y, Burns RD, Brusseau TA, Byun W. Effectiveness and feasibility of family and peer challenge intervention with wearable technology on physical activity among middle schoolers in Utah: a pilot trial. Prev Med Rep 2025 Jun;54:103095. [doi: 10.1016/j.pmedr.2025.103095] [Medline: 40469248]
83. Bicki AC, Seth D, McCulloch CE, Lin F, Ku E. Use of activity trackers to improve blood pressure in young people at risk for cardiovascular disease: a pilot randomized controlled trial. Pediatr Nephrol 2024 Aug;39(8):2467-2474. [doi: 10.1007/s00467-024-06340-6] [Medline: 38503990]
84. Friedman DN, Lavery JA, Moskowitz CS, et al. Feasibility and acceptability of a parent–child intervention to improve STEP count in childhood cancer survivors exposed to cardiotoxic therapy: the STEP UP for FAMILIES study. Pediatr Blood Cancer 2024 Dec;71(12):e31320. [doi: 10.1002/pbc.31320] [Medline: 39285785]
85. Gómez-Cuesta N, Mateo-Orcajada A, Meroño L, Abenza-Cano L, Vaquero-Cristóbal R. A mobile app-based intervention improves anthropometry, body composition and fitness, regardless of previous active-inactive status: a randomized controlled trial. Front Public Health 2024;12:1380621. [doi: 10.3389/fpubh.2024.1380621] [Medline: 39193194]
86. Ahmadi S, Gilbert JA, Marcotte M, de Fátima Guimarães R, Mathieu ME. Exercise-related self-perception, physical activity and intention to in-person and virtual activities among adolescent girls. Sports Med Health Sci 2025 Mar;7(2):109-115. [doi: 10.1016/j.smhs.2024.03.007] [Medline: 39811402]
87. Braga BC, Nguyen PH, Tran LM, et al. Feasibility of using an artificial intelligence-based telephone application for dietary assessment and nudging to improve the quality of food choices of female adolescents in Vietnam: evidence from a randomized pilot study. Curr Dev Nutr 2024 Jun;8(6):102063. [doi: 10.1016/j.cdnut.2023.102063] [Medline: 38817706]
88. Gilliland J, Bowman D, Cappuccitti S, et al. SmartAPPetite For Youth: pilot and feasibility study of an adolescent smartphone nutrition intervention. Pilot Feasibility Stud 2026 Feb 6;12(1):39. [doi: 10.1186/s40814-026-01767-0] [Medline: 41652603]




XSL•FO
RenderX

89. Moore SE, Rooney C, Neville CE, et al. The feasibility and acceptability of a rewards system based on food purchasing behaviour in secondary school cashless canteens: the Eat4Treats (E4T) cluster feasibility, non-randomised, controlled intervention study. Pilot Feasibility Stud 2024 Jan 9;10(1):4. [doi: 10.1186/s40814-023-01436-6] [Medline: 38195663]
90. Mays D, Macisco JM, Hawkins KB, et al. A multilevel intervention in pediatric primary care for youth tobacco control: outcomes of implementing an Ask, Advise, and Connect model. Transl Behav Med 2024 Mar 21;14(4):241-248. [doi: 10.1093/tbm/ibae002] [Medline: 38330454]
91. Lustig E, DiFiore G, Powell M, Leone FT, Schnoll R, Jenssen BP. Text-message outreach to connect adolescents and young adults to nicotine treatment: a pilot study in pediatric primary care. Acad Pediatr 2026 Jul;26(5):103318. [doi: 10.1016/j.acap.2026.103318] [Medline: 41967674]
92. Ling PM, Lyu JC, Lisha NE, Nguyen N, Gribben V, Ramo DE. Social media support groups for youth nicotine vaping cessation: an RCT. Am J Prev Med 2026 Aug;71(2):108314. [doi: 10.1016/j.amepre.2026.108314] [Medline: 41724253]
93. Sadasivam RS, Stevens EM, Wijesundara J, et al. Feasibility and engagement of a peer-driven mobile intervention for adolescent e-cigarette cessation: cluster randomized pilot study. JMIR Form Res 2026 Mar 17;10:e79667. [doi: 10.2196/79667] [Medline: 41843855]
94. Graham AL, Cha S, Jacobs MA, et al. A vaping cessation text message program for adolescent e-cigarette users: a randomized clinical trial. JAMA 2024 Sep 3;332(9):713-721. [doi: 10.1001/jama.2024.11057] [Medline: 39110436]
95. Fernández-León P, Lima-Serrano M, Martínez-Montilla JM. Usability evaluation of Alerta Alcohol 2.0: an eHealth game to prevent adolescent alcohol consumption. J Public Health 2026 Jun 1;48(2):477-487. [doi: 10.1093/pubmed/fdag022]
96. Wall H, Hansson H, Zetterlind U, Kvillemo P, Elgán TH. Effectiveness of a web-based individual coping and alcohol intervention program for children of parents with alcohol use problems: randomized controlled trial. J Med Internet Res 2024 Apr 10;26:e52118. [doi: 10.2196/52118] [Medline: 38598286]
97. Wang-Schweig M, Miller BA, Byrnes HF, Rogers V. Parent and teen engagement in Smart Choices 4 Teens to address teen alcohol use and relationships. J Child Fam Stud 2025 Jun;34(6):1417-1432. [doi: 10.1007/s10826-025-03074-3]
98. Berdiyarova S, Rakhmatova M, Uzbekova N, et al. Digital therapeutics and gamified mobile applications for behavioral modification in pediatric hypertension. Rev Latinoam Hipertens 2026;21:66-72. [doi: 10.5281/zenodo.18626596]
99. Chen YH, Tzeng YF, Gau BS. Mobile health-based health literacy weight management intervention using smart devices for adolescents: a convergent mixed-methods pilot study. Nurs Health Sci 2024 Dec;26(4):e13172. [doi: 10.1111/nhs.13172] [Medline: 39428322]
100. Oei K, Choi EE, Bar-Dayan A, et al. An image-recognition dietary assessment app for adolescents with obesity: pilot randomized controlled trial. JMIR Form Res 2024 Dec 2;8:e58682. [doi: 10.2196/58682] [Medline: 39621405]
101. Audi AS, Fernandes AE, Coelho GDM, et al. Effect of a mobile phone application for dietary self-monitoring on obesity in adolescents: a pilot randomized controlled trial. Eur J Clin Nutr 2026 Jan;80(1):96-103. [doi: 10.1038/s41430-025-01676-0] [Medline: 41193827]
102. Ramalho SM, Saint-Maurice PF, Silva D, Mansilha HF, Conceição E. Feasibility and effectiveness of a social network-based intervention for adolescents undergoing weight loss treatment: a randomized controlled trial. Nutrients 2025 Aug 8;17(16):2586. [doi: 10.3390/nu17162586] [Medline: 40871614]
103. O'Dean S, Sunderland M, Newton N, et al. The Health4Life e-health intervention for modifying lifestyle risk behaviours of adolescents: secondary outcomes of a cluster randomised controlled trial. Med J Aust 2024 May 6;220(8):417-424. [doi: 10.5694/mja2.52279] [Medline: 38613175]
104. Henry N, Montgomery E, Ghosh P, et al. Caregiver inclusion influence on adolescent acceptance and engagement of an mHealth app, a randomized controlled trial. Mhealth 2025;11:57. [doi: 10.21037/mhealth-24-97] [Medline: 41211016]
105. Bjerregaard AA, Zoughbie DE, Hansen JV, et al. An SMS chatbot digital educational program to increase healthy eating behaviors in adolescence: a multifactorial randomized controlled trial among 7,890 participants in the Danish National Birth Cohort. PLoS Med 2024 Jun;21(6):e1004383. [doi: 10.1371/journal.pmed.1004383] [Medline: 38875292]
106. Mateo-Orcajada A, Ponce-Ramírez CM, Abenza-Cano L, Vaquero-Cristóbal R. Effects of 10 weeks of walking with mobile step-tracking apps on body composition, fitness, and psychological state in adolescents who are overweight and obese: randomized controlled trial. J Med Internet Res 2024 Dec 10;26:e55243. [doi: 10.2196/55243] [Medline: 39656510]
107. Roche M, McIntyre A, Oliver C, et al. How can we better engage female athletes? A novel approach to health and performance education in adolescent athletes. BMJ Open Sport Exerc Med 2024;10(3):e001901. [doi: 10.1136/bmjsem-2024-001901] [Medline: 39290377]
108. Ponce-Ramírez CM, Mateo-Orcajada A, Abenza-Cano L, Vaquero-Cristóbal R. Assessment of step tracker mobile applications for the promotion of physical activity by adolescents based on their weight status. Int J Telemed Appl 2024;2024:8038334. [doi: 10.1155/2024/8038334] [Medline: 39381367]
109. Seiterö A, Henriksson P, Thomas K, et al. Effectiveness of a mobile phone-delivered multiple health behavior change intervention (LIFE4YOUth) in adolescents: randomized controlled trial. J Med Internet Res 2025 Apr 22;27:e69425. [doi: 10.2196/69425] [Medline: 40262133]
110. Egan L, Gardner LA, Newton NC, Champion KE. Process evaluation of the digital Health4Life intervention among a sample of disadvantaged adolescents and teachers. Health Promot Int 2024 Dec 1;39(6):1-10. [doi: 10.1093/heapro/daae170] [Medline: 39612512]




XSL•FO
RenderX

111. Hartung J, Knapp G. On tests of the overall treatment effect in meta-analysis with normally distributed responses. Stat Med 2001 Jun 30;20(12):1771-1782. [doi: 10.1002/sim.791] [Medline: 11406840]
112. Lin L, Chu H. Meta-analysis of proportions using generalized linear mixed models. Epidemiology 2020 Sep;31(5):713-717. [doi: 10.1097/EDE.0000000000001232] [Medline: 32657954]
113. Stijnen T, Hamza TH, Özdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. Statist Med 2010 Dec 20;29(29):3046-3067. [doi: 10.1002/sim.4040]
114. Van den Noortgate W, López-López JA, Marín-Martínez F, Sánchez-Meca J. Three-level meta-analysis of dependent effect sizes. Behav Res 2013 Jun;45(2):576-594. [doi: 10.3758/s13428-012-0261-6]
115. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634. [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
116. Campbell M, McKenzie JE, Sowden A, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ 2020 Jan 16;368:l6890. [doi: 10.1136/bmj.l6890] [Medline: 31948937]
117. R Core Team. R: a language and environment for statistical computing. : R Foundation for Statistical Computing; 2026 URL: https://www.R-project.org/ [accessed 2026-07-27]
118. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw 2010;36:1-48. [doi: 10.18637/jss.v036.i03]
119. Howell CR, Krull KR, Partin R, et al. Randomized web-based physical activity intervention in adolescent survivors of childhood cancer. J Clin Oncol 2018 Mar 1;36(7_suppl):102-102. [doi: 10.1200/JCO.2018.36.7_suppl.102]
120. Jago R, Baranowski T, Baranowski JC, et al. Fit for Life Boy Scout badge: outcome evaluation of a troop and Internet intervention. Prev Med 2006 Mar;42(3):181-187. [doi: 10.1016/j.ypmed.2005.12.010] [Medline: 16458955]
121. Klausen SH, Andersen LL, Søndergaard L, et al. Effects of eHealth physical activity encouragement in adolescents with complex congenital heart disease: the PReVaiL randomized clinical trial. Int J Cardiol 2016 Oct 15;221:1100-1106. [doi: 10.1016/j.ijcard.2016.07.092] [Medline: 27448540]
122. Lau EY, Lau PWC, Chung PK, Ransdell LB, Archer E. Evaluation of an Internet-short message service-based intervention for promoting physical activity in Hong Kong Chinese adolescent school children: a pilot study. Cyberpsychol Behav Soc Netw 2012 Aug;15(8):425-434. [doi: 10.1089/cyber.2012.0161] [Medline: 22897473]
123. Lelieveld O, Armbrust W, Geertzen JHB, et al. Promoting physical activity in children with juvenile idiopathic arthritis through an internet‐based program: results of a pilot randomized controlled trial. Arthritis Care Res 2010 May;62(5):697-703 [FREE Full text] [doi: 10.1002/acr.20085]
124. Meyer M, Brudy L, Fuertes-Moure A, et al. E-Health exercise intervention for pediatric patients with congenital heart disease: a randomized controlled trial. J Pediatr 2021 Jun;233:163-168. [doi: 10.1016/j.jpeds.2021.01.058] [Medline: 33516681]
125. Slootmaker SM, Chinapaw MJM, Seidell JC, van Mechelen W, Schuit AJ. Accelerometers and Internet for physical activity promotion in youth? Feasibility and effectiveness of a minimal intervention [ISRCTN93896459]. Prev Med 2010 Jul;51(1):31-36. [doi: 10.1016/j.ypmed.2010.03.015] [Medline: 20380847]
126. Pedersen S, Grønhøj A, Thøgersen J. Texting your way to healthier eating? Effects of participating in a feedback intervention using text messaging on adolescents' fruit and vegetable intake. Health Educ Res 2016 Apr;31(2):171-184. [doi: 10.1093/her/cyv104] [Medline: 26850061]
127. Doyle AC, Goldschmidt A, Huang C, Winzelberg AJ, Taylor CB, Wilfley DE. Reduction of overweight and eating disorder symptoms via the internet in adolescents: a randomized controlled trial. J Adolesc Health 2008 Aug;43(2):172-179. [doi: 10.1016/j.jadohealth.2008.01.011] [Medline: 18639791]
128. Likhitweerawong N, Boonchooduang N, Kittisakmontri K, Chonchaiya W, Louthrenoo O. Effectiveness of mobile application on changing weight, healthy eating habits, and quality of life in children and adolescents with obesity: a randomized controlled trial. BMC Pediatr 2021 Nov 10;21(1):499. [doi: 10.1186/s12887-021-02980-x] [Medline: 34758761]
129. Lopez KE, Salvy SJ, Fink C, et al. Executive functioning, depressive symptoms, and intervention engagement in a sample of adolescents enrolled in a weight management program. Child Obes 2021 Jun;17(4):281-290. [doi: 10.1089/chi.2020.0334] [Medline: 33826861]
130. Vidmar AP, Pretlow R, Borzutzky C, et al. An addiction model-based mobile health weight loss intervention in adolescents with obesity. Pediatr Obes 2019 Feb;14(2):e12464. [doi: 10.1111/ijpo.12464] [Medline: 30117309]
131. Aveyard P, Markham WA, Almond J, Lancashire E, Cheng KK. The risk of smoking in relation to engagement with a school-based smoking intervention. Soc Sci Med 2003 Feb;56(4):869-882. [doi: 10.1016/s0277-9536(02)00088-6] [Medline: 12560019]
132. Mason M, Mennis J, Way T, et al. Text message delivered peer network counseling for adolescent smokers: a randomized controlled trial. J Prim Prev 2016 Oct;37(5):403-420. [doi: 10.1007/s10935-016-0439-2] [Medline: 27388626]
133. Mays D, Hawkins KB, Bredfeldt C, Wolf H, Tercyak KP. The effects of framed messages for engaging adolescents with online smoking prevention interventions. Transl Behav Med 2017 Jun;7(2):196-203. [doi: 10.1007/s13142-017-0481-5] [Medline: 28290144]




XSL•FO
RenderX

134. Pallonen UE, Velicer WF, Prochaska JO, et al. Computer-based smoking cessation interventions in adolescents: description, feasibility, and six-month follow-up findings. Subst Use Misuse 1998 Mar;33(4):935-965. [doi: 10.3109/10826089809056250] [Medline: 9548631]
135. Patten CA, Croghan IT, Meis TM, et al. Randomized clinical trial of an Internet-based versus brief office intervention for adolescent smoking cessation. Patient Educ Couns 2006 Dec;64(1-3):249-258. [doi: 10.1016/j.pec.2006.03.001] [Medline: 16616449]
136. Pbert L, Druker S, Crawford S, et al. Feasibility of a smartphone app with mindfulness training for adolescent smoking cessation: Craving to Quit (C2Q)-Teen. Mindfulness (N Y) 2020 Mar;11(3):720-733. [doi: 10.1007/s12671-019-01273-w] [Medline: 33343761]
137. Ho FK, Tung KTS, Wong RS, et al. An internet quiz game intervention for adolescent alcohol drinking: a clustered RCT. Pediatrics 2021 Dec 1;148(6):e2021051005. [doi: 10.1542/peds.2021-051005] [Medline: 34841427]
138. Newton NC, Stapinski LA, Slade T, et al. The 7-year effectiveness of school-based alcohol use prevention from adolescence to early adulthood: a randomized controlled trial of universal, selective, and combined interventions. J Am Acad Child Adolesc Psychiatry 2022 Apr;61(4):520-532. [doi: 10.1016/j.jaac.2021.10.023] [Medline: 34823025]
139. Schinke SP, Schwinn TM, Fang L. Longitudinal outcomes of an alcohol abuse prevention program for urban adolescents. J Adolesc Health 2010 May;46(5):451-457. [doi: 10.1016/j.jadohealth.2009.11.208] [Medline: 20413081]
140. Vargas-Martínez AM, Trapero-Bertran M, Lima-Serrano M, Anokye N, Pokhrel S, Mora T. Measuring the effects on quality of life and alcohol consumption of a program to reduce binge drinking in Spanish adolescents. Drug Alcohol Depend 2019 Dec 1;205:107597. [doi: 10.1016/j.drugalcdep.2019.107597] [Medline: 31590139]
141. Abroms LC, Fagan P, Eisenberg ME, Lee HSH, Remba N, Sorensen G. The STRENGTH Ezine: an application of e-mail for health promotion in adolescent girls. Am J Health Promot 2004;19(1):28-32. [doi: 10.4278/0890-1171-19.1.28] [Medline: 15460098]
142. Majumdar D, Koch PA, Lee H, Contento IR, Islas-Ramos ADL, Fu D. “Creature-101”: a serious game to promote energy balance-related behaviors among middle school adolescents. Games Health J 2013 Oct;2(5):280-290. [doi: 10.1089/g4h.2013.0045] [Medline: 24761326]
143. Schwinn TM, Fang L, Hopkins J, Pacheco AR. Longitudinal outcomes of a smartphone application to prevent drug use among Hispanic youth. J Stud Alcohol Drugs 2021 Sep;82(5):668-677. [doi: 10.15288/jsad.2021.82.668] [Medline: 34546914]
144. Mateo-Orcajada A, Vaquero-Cristóbal R, Abenza-Cano L. Importance of training volume through the use of step trackers apps promoted from the subject of physical education to change body composition, physical fitness and physical activity in adolescents and the influence of gender. Physiol Behav 2024 Jan 1;273:114402. [doi: 10.1016/j.physbeh.2023.114402] [Medline: 37935300]
145. Paz Castro R, Haug S, Debelak R, Jakob R, Kowatsch T, Schaub MP. Engagement with a mobile phone-based life skills intervention for adolescents and its association with participant characteristics and outcomes: tree-based analysis. J Med Internet Res 2022 Jan 19;24(1):e28638. [doi: 10.2196/28638] [Medline: 35044309]
146. Whittemore R, Jeon S, Grey M. An internet obesity prevention program for adolescents. J Adolesc Health 2013 Apr;52(4):439-447. [doi: 10.1016/j.jadohealth.2012.07.014] [Medline: 23299003]
147. Parker MN, Bloomer BF, Stout JD, et al. A pilot randomized control trial testing a smartphone-delivered food attention retraining program in adolescent girls with overweight or obesity. Nutrients 2024 Oct 12;16(20):3456. [doi: 10.3390/nu16203456] [Medline: 39458453]
148. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52. [doi: 10.2196/jmir.1772] [Medline: 21821503]
149. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68] [Medline: 11392867]
150. Johnson D, Deterding S, Kuhn KA, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106. [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
151. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". Presented at: MindTrek ’11: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments; Sep 28-30, 2011; Tampere, Finland p. 9-15. [doi: 10.1145/2181037.2181040]
152. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30. [doi: 10.2196/jmir.1602] [Medline: 21393123]
153. Schuster R, Leitner I, Carlbring P, Laireiter AR. Exploring blended group interventions for depression: randomised controlled feasibility study of a blended computer- and multimedia-supported psychoeducational group intervention for adults with depressive symptoms. Internet Interv 2017 Jun;8:63-71. [doi: 10.1016/j.invent.2017.04.001] [Medline: 30135830]
154. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
155. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]




XSL•FO
RenderX

156. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
157. Woolley MG, Klimczak KS, Davis CH, Levin ME. Predictors of adherence to a publicly available self-guided digital mental health intervention. Cogn Behav Ther 2024 Nov;53(6):577-591. [doi: 10.1080/16506073.2024.2341807] [Medline: 38619511]
158. Szinay D, Forbes CC, Busse H, DeSmet A, Smit ES, König LM. Is the uptake, engagement, and effectiveness of exclusively mobile interventions for the promotion of weight-related behaviors equal for all? A systematic review. Obes Rev 2023 Mar;24(3):e13542. [doi: 10.1111/obr.13542] [Medline: 36625062]
159. Western MJ, Smit ES, Gültzow T, et al. Bridging the digital health divide: a narrative review of the causes, implications, and solutions for digital health inequalities. Health Psychol Behav Med 2025;13(1):2493139. [doi: 10.1080/21642850.2025.2493139] [Medline: 40276490]
160. Pham Thi TD, Duong NT. Intrinsic motivations in health and fitness app engagement: a mediation model of entertainment. Digit Health 2025;11:20552076251326151. [doi: 10.1177/20552076251326151] [Medline: 40093708]
161. Melo GLR, Santo RE, Mas Clavel E, et al. Digital dietary interventions for healthy adolescents: a systematic review of behavior change techniques, engagement strategies, and adherence. Clin Nutr 2025 Feb;45:176-192. [doi: 10.1016/j.clnu.2025.01.012] [Medline: 39827503]
162. Li Z, Lu F, Wu J, et al. Usability and effectiveness of eHealth and mHealth interventions that support self-management and health care transition in adolescents and young adults with chronic disease: systematic review. J Med Internet Res 2024 Nov 26;26:e56556. [doi: 10.2196/56556] [Medline: 39589770]
163. Ho TQA, Le LKD, Engel L, et al. Barriers to and facilitators of user engagement with web-based mental health interventions in young people: a systematic review. Eur Child Adolesc Psychiatry 2025 Jan;34(1):83-100. [doi: 10.1007/s00787-024-02386-x]
164. Wanniarachchi VU, Greenhalgh C, Warren J. Adolescents' and youths' perceived barriers and facilitators to engaging with digital mental health interventions for depression and anxiety: a scoping review. Internet Interv 2025 Dec;42:100884. [doi: 10.1016/j.invent.2025.100884] [Medline: 41210520]
165. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
166. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919. [doi: 10.1136/bmj.i4919] [Medline: 27733354]
167. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ 2010 Feb 5;340:c221. [doi: 10.1136/bmj.c221] [Medline: 20139215]


## Abbreviations

**PEO:** Population-Exposure-Outcome
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**RCT:** randomized controlled trial
**RoB 2:** Cochrane Risk of Bias tool for randomized trials
**ROBINS-I:** Risk Of Bias In Non-randomized Studies of Interventions
**SUNRISE:** Sustainable Interventions and Healthy Behaviours for Adolescent Primary Prevention of Cancer with Digital Tools
**SWiM:** Synthesis Without Meta-analysis
**WHO:** World Health Organization


*Edited by M Balcarras; submitted 25.Sep.2025; peer-reviewed by A DeSmet, A Maba, M Lima-Serrano; revised version received 03.Aug.2026; accepted 06.Aug.2026; published 30.Sep.2026.*

*Please cite as:*
*Boumparis N, Studhalter O, de Riedmatten P, Yücel ID, Koutra K, Champion K, Molina-Barceló A, de Pablo-Pardo T, Kondylakis H, Schaub MP, Triantafyllidis A, Haug S*
*Factors Influencing Adherence to Digital Lifestyle Interventions for Adolescents: Systematic Review and Meta-Analysis of Attrition*
*Interact J Med Res 2026;15:e84822*
*URL: https://www.i-jmr.org/2026/1/e84822*
*doi:10.2196/84822*


© Nikolaos Boumparis, Olivia Studhalter, Philippe de Riedmatten, Inci Derya Yücel, Kleio Koutra, Katrina Champion, Ana Molina-Barceló, Teresa de Pablo-Pardo, Haridimos Kondylakis, Michael Patrick Schaub, Andreas Triantafyllidis, Severin Haug. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 30.Sep.2026. This is an open-access



XSL•FO
RenderX

article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Design, Facilitation, and Evaluation of Tabletop Exercises for Prehospital Mass Casualty Preparedness: Scoping Review

Paurnami Prashanth[1], MBBS; Ali AlRahma[1], MD; Dayol Narayanan[2], BSN, MBA; Abu Omayer[3], MD; Azza Yousif[1], MBBS, MSc; Ives Hubloue[4,5], PhD; Nabil Zary[3], PhD

[1]Emergency Department, Dubai Health, Umm Hurair Second, Bur Dubai, Dubai, United Arab Emirates

[2]Dubai Corporation for Ambulance Services, Dubai, United Arab Emirates

[3]Institute of Learning, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai Health, Dubai, United Arab Emirates

[4]Emergency Department, UZ Brussel University Hospital, Brussels, Belgium

[5]Research Group on Emergency and Disaster Medicine (ReGEDiM) Vrije Universiteit Brussel, Brussels, Belgium

**Corresponding Author:**
Azza Yousif, MBBS, MSc
Emergency Department, Dubai Health, Umm Hurair Second, Bur Dubai, Dubai, United Arab Emirates

## Abstract

**Background:** Tabletop exercises (TTXs) are commonly used to enhance prehospital readiness for mass casualty incidents (MCIs). However, evidence on their design, facilitation, and evaluation remains scattered. TTXs simulate organized interactions at 3 levels: among individual responders and response protocols, within interdisciplinary teams, and across organizations and systems. While existing reviews cover tabletop simulation generally, they do not specifically focus on the prehospital MCI interface or assess whether evaluation methods match the exercise objectives.

**Objective:** This scoping review aimed to explore how TTXs are designed, facilitated, and evaluated in prehospital MCI preparedness. It classified outcomes using the Kirkpatrick Evaluation Model and analyzed how evaluation methods align with the purpose of each exercise.

**Methods:** We performed a scoping review following Arksey and O'Malley's framework, with enhancements from Levac et al, along with PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) and PRISMA-S (PRISMA Extension for Reporting Literature Searches) guidelines. Searches across PubMed, Embase, Scopus, PsycINFO, CINAHL, the Cochrane Library, ClinicalTrials.gov, Google Scholar, and references were limited to English peer-reviewed and gray literature published from 2015 to May 2026. Eligible studies focused on TTXs related to prehospital disaster or MCI preparedness and reported measurable educational, clinical, or system outcomes. The review protocol was registered beforehand on Protocols.io. Two reviewers (PP and AA) independently screened records, extracted data, and classified outcomes using Kirkpatrick levels, with an added level 2+ (Applied Learning) category for structured performance evaluation during exercises.

**Results:** Thirteen studies from 9 countries were included. They were categorized into a preliminary 3-tier typology called the TTX Design Spectrum, representing increasing levels of interaction: algorithm-rehearsal exercises (n=2) focused on individual responders' interaction with triage protocols, scenario-based decision-training exercises (n=7) aimed at multidisciplinary team decision-making under realistic conditions, and systems integration exercises (n=4) centered on interagency coordination and system-level preparedness. Facilitation varied by exercise purpose, from standardized assessment-focused approaches to expert-led and multidisciplinary facilitation. Evaluation primarily targeted Kirkpatrick level 1: Reaction (n=10), level 2: Learning (n=10), and level 2+ (Applied Learning) (n=8), with fewer studies examining level 3: Behavior (n=2) or level 4: Results (n=3). Operational frameworks were reported more consistently than formal educational design or assessment frameworks. Additionally, natural disaster scenarios and evidence from resource-limited settings were underrepresented.

**Conclusions:** TTXs for prehospital MCI preparedness should be viewed as a collection of related exercise types rather than a single, uniform intervention. This review introduces an initial typology called the TTX Design Spectrum, along with the level 2+ (Applied Learning) classification, which operationalizes the distinction between in-training performance and behavioral transfer. These tools aim to help align exercise purpose, facilitation, and evaluation strategies. Future research should focus on validating this typology, enhancing follow-up assessments of behavioral transfer, and adapting TTX design for natural disaster and resource-limited settings.

***(Interact J Med Res 2026;15:e96228)*** doi:10.2196/96228



XSL•FO
RenderX

**KEYWORDS**

tabletop exercise; mass casualty incident; prehospital; scoping review; Kirkpatrick Evaluation Model; simulation-based education; disaster preparedness; applied learning; training transfer; exercise design

## Introduction

Mass casualty incidents (MCIs) and disasters continue to challenge health systems because they demand quick coordination, prioritization, and adaptation amid uncertainty, time constraints, and limited resources. The Sendai Framework for Disaster Risk Reduction emphasizes the importance of preparedness for effective response in modern disaster management [1]. Prehospital readiness involves personnel and systems managing triage, incident command, scene coordination, interagency communication, resource management, and treatment prioritization. This preparedness significantly impacts outcomes when systems are overwhelmed. As a result, health care simulation use has increased [2,3], and recent tabletop and disaster education research highlights incident command, triage, surge management, communication, and resource allocation as essential skills that require practice before real emergencies [4-6].

Tabletop exercises (TTXs) stand out among simulation methods because they are structured, discussion-driven activities where participants can rehearse plans, roles, communication, and operational choices without the logistical challenges of full-scale or functional exercises [5,7]. They are also versatile, being customizable to an organization's specific risks with adaptable goals, scope, and difficulty levels [2,5], making them a popular tool for improving disaster and MCI preparedness [4,8-10].

The existing literature indicates that TTXs are not a uniform type of intervention. Some primarily concentrate on practicing triage algorithms, while others focus on team decision-making, interagency communication, or organizational coordination [2,4-6]. This variation is expected because exercises serve different educational and operational goals. The challenge lies in the fact that the term "tabletop exercise" is often used to describe quite different activities. Without a common framework for describing the purpose of each exercise, educators and planners struggle to determine whether an evaluation method fits the exercise's objectives and find it difficult to compare findings across studies that use the same term for different types of activities.

At their core, TTXs are organized interactions involving individual responders, the protocols they follow, professionals from various disciplines, and different organizations, agencies, and systems. The specific level of interaction an exercise aims to rehearse determines who participates, how it is facilitated, and what metrics should be used for evaluation. Consequently, a typology of exercise purposes also reflects the types of interactions being trained. This shared vocabulary helps determine whether an evaluation method is appropriate for a particular exercise and allows for comparison of evidence across studies.

Recent review literature highlights the need for a more targeted synthesis. Frégeau et al [7] conducted a comprehensive scoping review of tabletop simulations in medical emergencies, analyzing 70 studies across various settings, specialties, formats, and learner groups. They found that reported outcomes mainly focused on the Kirkpatrick reaction and learning levels [7]. Emaliyawati et al [8] reviewed 12 tabletop disaster exercise studies involving health care workers and students, concluding that TTXs enhance knowledge, attitudes, preparedness, confidence, and performance. Although these reviews are valuable, they address TTXs broadly, rather than focusing specifically on the prehospital interface. They do not analyze how prehospital MCI competencies are distributed across different exercise designs, nor do they systematically link exercise purpose with facilitation style and evaluation depth. As far as we know, no published review has specifically synthesized TTXs for prehospital MCI preparedness while examining the alignment among exercise design, facilitation, and evaluation [7,8].

This scoping review aimed to map the existing literature and clarify how TTXs function within prehospital MCI preparedness, considering them as a group of related but distinct interventions. The objectives included systematically investigating how TTXs have been used to evaluate or enhance prehospital readiness for MCIs; classifying reported outcomes according to the Kirkpatrick Evaluation Model, including structured performance within exercises; describing key design, delivery, and facilitation features along with the competencies they target; and pinpointing gaps in the evidence. The review seeks to develop a preliminary typology called the TTX Design Spectrum, which aims to guide purpose-aligned design, facilitation, and evaluation in prehospital disaster preparedness.

## Methods

### Study Design and Methodological Framework

#### *Overview*

This scoping review, part of the Mass-Casualty Incident, Prehospital Emergency Response project, systematically maps evidence related to TTXs for prehospital disaster preparedness. Our methodology follows Arksey and O'Malley's [11] framework, with modifications by Levac et al [12], and adheres to PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) reporting standards [13]. The protocol was registered on Protocols.io (July 3, 2025) [14] and published as a preprint [15], ensuring transparency and reproducibility. Stakeholder consultation, recommended by Levac et al [12] for knowledge translation, was not carried out because the study primarily focused on evidence mapping for research synthesis rather than developing practice guidelines.

#### *Stage 1: Identifying the Research Question*

This scoping review was based on 2 main research questions and 3 subsidiary questions. The primary questions explored how TTXs are used to assess or enhance prehospital preparedness for MCIs, and which Kirkpatrick evaluation levels



XSL•FO
RenderX

(Reaction, Learning, Behavior, and Results) are most frequently reported in prehospital MCI TTX studies. The secondary questions investigated the key design and delivery features of these TTXs, the competencies and operational areas they target, and identified gaps in evidence concerning evaluation methods and their effectiveness.

### *Stage 2: Identifying Relevant Studies*

Search reporting followed the PRISMA-S (PRISMA Extension for Reporting Literature Searches) extension for literature search reporting [16], along with PRISMA-ScR. A completed PRISMA-ScR checklist is included in Checklist 1. We searched several multidisciplinary databases, including PubMed, Embase, Scopus, PsycINFO (via APA PsycNet), CINAHL, Cochrane Library, and ClinicalTrials.gov, to cover biomedical, educational, and allied health research. Each database and registry was searched individually; no multidatabase search was performed simultaneously. We also reviewed Google Scholar for gray literature and manually screened reference lists of included studies and relevant reviews. No contact was made with study authors, experts, manufacturers, or external contacts to gather additional records or data. Our search methods included only database searches, Google Scholar, and manual reference list screening.

On June 10, 2025, we established a detailed search strategy based on the Population-Concept-Context framework (Table 1), which was subsequently updated on May 15, 2026. The Population targeted prehospital health care workers such as paramedics, emergency medical technicians, emergency physicians, and nurses, as well as emergency system administrators involved in MCI responses. The Concept focused on TTXs used for training and preparedness, with outcomes measurable at the educational, clinical, or system levels. The Context covered prehospital systems that respond to mass-casualty incidents and disasters across various geographic areas.

**Table .** Search strategy overview using the Population-Concept-Context framework. Representative controlled vocabulary and free-text terms are displayed. Detailed database-specific search strings are available in Multimedia Appendix 1.

| Category | Keywords | Search strategy |
|---|---|---|
| Population | “Emergency Medical Services” [MeSH]; “Emergency Medical Technicians” [MeSH]; “Allied Health Personnel” [MeSH] | paramedic*; “first responder*”; “first responder”; “first responders”; ambulance*; “emergency medical service*”; “emergency medical service”; “emergency medical services”; EMS[a] |
| Concept | “Simulation Training” [MeSH]; “Patient Simulation” [MeSH]; “Computer Simulation” [MeSH]; “simulation training”; “tabletop simulation”; “tabletop exercise” | tabletop exercise*; table-top exercise*; tabletop simulation*; tabletop drill*; discussion-based exercise*; scenario-based simulation*; preparedness exercise*; board game*; simulation game*; gamified exercise; paper-based exercise |
| Context | “Mass Casualty Incidents” [MeSH]; “Disaster Planning” [MeSH]; “Emergency Preparedness” [MeSH]; “Triage” [MeSH]; “mass disaster”; “field triage”; “incident command”; “prehospital care” | emergency preparedness; disaster preparedness; mass casualty incident*; MCI[b]; field triage; incident command; prehospital care; disaster simulation; response coordination; crisis; disaster response; multi-casualty; multiple casualty |

[a]EMS: emergency medical services.

[b]MCI: mass casualty incident.

We used a combination of controlled vocabulary (MeSH) and free-text terms with Boolean operators in search strings tailored to each database, refining them iteratively to improve sensitivity and specificity (Table 1). An information specialist or librarian peer-reviewed the search strategy following the Peer Review of Electronic Search Strategies criteria before final implementation [17]. No published search filters were used, nor were previous review strategies reused or adapted; instead, prior reviews helped identify relevant references through citation screening. Details about platforms and complete database-specific search strings, including the full PubMed strategy, are available in Multimedia Appendix 1.

### *Stage 3: Study Selection*

Retrieved records from various sources, including citation searches and gray literature, were imported into Covidence [18], where they were pooled and deduplicated before screening. Two reviewers, PP and AA, independently examined all records through a 2-stage process: first screening titles and abstracts and then reviewing full texts. Any disagreements were resolved through discussion or, if necessary, by consulting a third reviewer. These disagreements were systematically recorded in Covidence and addressed through structured discussions with documented reasons.

## Eligibility Criteria

### *Overview*

We included studies that assessed TTXs in prehospital disaster or MCI preparedness and reported measurable educational, clinical, or system outcomes. These were peer-reviewed papers or gray literature published in English between January 2015 and May 2026. For this review, “prehospital” was mainly defined by the medical phase and operational functions of MCI or disaster response, rather than solely by participant role or physical setting. This encompassed EMS (emergency medical services) care, field triage, scene coordination, casualty distribution, transport decisions, and the transition from prehospital to hospital care. Eligible study designs included quantitative, qualitative, mixed methods, quasi-experimental,



XSL•FO
RenderX

pre-post, randomized, observational, descriptive, pilot, feasibility, and implementation studies. We excluded systematic reviews, narrative reviews, editorials, commentaries, opinion pieces, conference abstracts without full texts, studies lacking full-text access, non-English studies, studies published before 2015, and studies where TTXs were not the primary focus. Additionally, studies focused on nonmedical TTX applications, such as military, corporate, cybersecurity, or other non–health care settings, were excluded unless they included a relevant medical prehospital or MCI response component. Studies involving hospital personnel were included only if they addressed prehospital medical functions, EMS coordination, transport, casualty distribution, or the hospital transition process.

We chose papers from 2015 onward because initial scoping indicated limited peer-reviewed literature before 2015 on TTX outcomes in prehospital MCI settings. Furthermore, reporting standards for simulation-based education improved significantly during this period [19], and the 2015 Sendai Framework for Disaster Risk Reduction [1] provided a current policy context for disaster preparedness research.

### *Stage 4: Data Charting*

Before starting the formal data extraction, we tested a standardized data-charting form with 2 reviewers (DN and AO) across 5 studies to verify that the extraction fields were clear, comprehensive, and consistently understood. During this pilot, reviewers independently extracted data, compared their entries, discussed any discrepancies, and refined the form before proceeding with full extraction. The finalized form included study details (ID, country, design, population, and sample size), TTX features (targeted competencies, duration, scenario design, and preparatory instruction), facilitation aspects, evaluation methods, Kirkpatrick levels addressed, and the underlying training, assessment, and operational frameworks.

After the data-charting form was finalized, 2 reviewers (DN and AO) independently extracted data from all included studies. They compared their extractions, resolving any discrepancies through discussion and consensus. If they could not reach an agreement, a third reviewer was consulted. Microsoft Excel was used to manage, review, and organize the extracted data.

Outcomes were categorized using the Kirkpatrick Evaluation Model [20,21]. Beyond the 4 standard levels, we added a level 2+ (Applied Learning) subcategory to differentiate structured performance assessments within the exercise setting from knowledge acquisition and actual behavioral transfer in real-world contexts. This subcategory helps operationalize, for TTX outcome mapping, distinctions outlined in earlier frameworks: the updated (“New World”) Kirkpatrick model’s focus on skills shown during training [22], Miller’s “shows how” level of assessment [23], and translational science models of simulation-based education outcomes (T1-T3), which differentiate in-simulation performance from later behavior and patient or system results [24]. Each outcome was classified independently; thus, a single study could contribute to multiple Kirkpatrick levels. Detailed definitions, decision rules, and examples for classification are available in Multimedia Appendix 2.

### *Stage 5: Collating, Summarizing, and Reporting the Results*

We used a combination of descriptive and narrative approaches appropriate for scoping reviews. Quantitative information was summarized using frequencies and counts, while qualitative and contextual data, such as exercise design, scenario features, facilitation processes, targeted skills, and evaluation techniques, were synthesized narratively. The extracted data were categorized into 7 domains: general study details, population traits, TTX features and elements, facilitator responsibilities, data collection methods, training results, and frameworks that inform training design, assessment, and operational practice.

Through iterative cross-study comparison, we discovered common patterns in exercise design, facilitation, and evaluation. To categorize these patterns, we created the TTX Design Spectrum, which classifies exercises into 3 tiers based on their primary purpose and scope: tier 1 focuses on algorithm rehearsal, tier 2 on scenario-based decision training, and tier 3 on systems integration testing. Facilitation methods and evaluation patterns were summarized descriptively rather than formalized into additional models. This typology was developed inductively during synthesis rather than being predefined; it is an initial organizational tool and not a validated framework. Definitions can be found in Multimedia Appendix 2.

### Protocol Deviations

This review was based on a prospectively registered protocol [14,15]. Two post hoc analytical refinements were introduced: the creation of the TTX Design Spectrum typology, which provides a descriptive overview of facilitation and evaluation patterns during stage 5, and the addition of the level 2+ (Applied Learning) subcategory in outcome classification during stage 4. Neither of these changes was specified in the protocol; both were developed through iterative comparison and synthesis of data. Importantly, there were no deviations that affected the search strategy or eligibility criteria.

### Ethical Considerations

This scoping review summarized publicly accessible, previously published literature without direct involvement of human participants. Therefore, ethical approval was not required. We adhered to established ethical research principles, such as transparent reporting, proper attribution and citation, and preventing plagiarism and duplicate publications.

## Results

### Literature Search and Study Selection

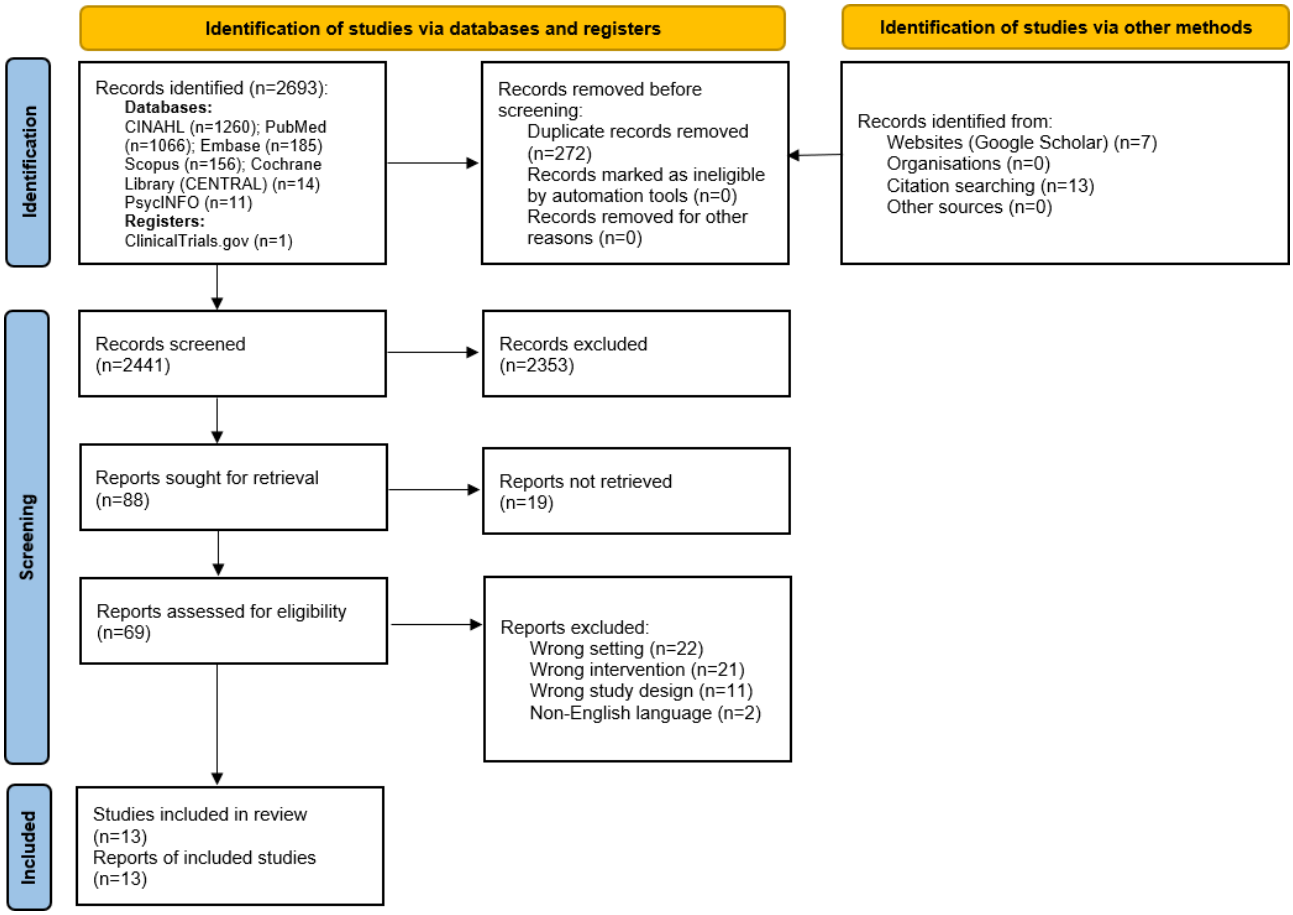
Our search identified 2693 records from databases and trial registers, along with 20 additional records from other sources. After removing 272 duplicates, we screened 2441 records by title and abstract, excluding 2353. We attempted to retrieve 88 full-text reports; of these, 19 were conference abstracts without full publications. Metadata for these 19 records (author, year, title, and source) is provided in Multimedia Appendix 3. The publication years and geographic origins of unretrieved records were similar to those of the retrieved records (Multimedia Appendix 3). Of the 69 full-text reports assessed for eligibility,



XSL•FO
RenderX

56 were excluded (wrong setting, n=22; wrong intervention, n=21; wrong study design, n=11; and non-English, n=2). Thirteen studies met the inclusion criteria and were included in the final analysis. Figure 1 illustrates the study selection process.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram for study selection. Records were identified through 6 databases, 1 register, and other methods. Of these, 13 studies met all inclusion criteria and were categorized into 3 tiers of the Tabletop Exercise Design Spectrum.




## Study Characteristics

Thirteen studies published between 2016 and 2026 were included, with most publications in 2020 (n=3). The studies took place in various countries: 4 in the United States, 2 in Thailand, and 1 each in the United Kingdom, Saudi Arabia, South Korea, Qatar, Spain, Sweden, and Japan. The research designs comprised 4 pre-post evaluations, 3 quasi-experimental studies, 2 reliability assessments, 2 mixed methods investigations, 1 randomized controlled trial, and 1 descriptive report. The sample sizes varied from 12 to 230 participants; 1 study did not specify the number of fellows involved (refer to Table 2, footnote).



XSL•FO
RenderX

**Table .** Characteristics of Included Studies (N=13).

| Study | Country | Design | Participants, n | Population | Duration | Scenarios |
|---|---|---|---|---|---|---|
| Tier 1: algorithm rehearsal (n=2) | | | | | | |
| Cheng et al (2022) [25] | United States | Reliability study | 107 | Prehospital EMS[a] professionals | 4 - 11 minutes per set | 1 (pediatric bus crash, 25 patients; 6 triage algorithms compared) |
| McGlynn et al (2020) [26] | United States | Reliability study | NR[b] | Pediatric emergency medicine fellows | Not reported | 3 (10-, 100-, and 1000-victim MCI[c] scenarios; 253 patient cases) |
| Tier 2: scenario-based decision training (n=7) | | | | | | |
| Kim et al (2021) [27] | South Korea | Pre-post | 40 | Physicians, nurses, and public health officers | 25 minutes + 10 minutes debrief | 1 (hydrofluoric acid explosion at chemical plant) |
| Phattharapornjaroen et al (2020) [28] | Thailand | Pre-post | 52 | Emergency physicians | 2-day course | 3 (building fire, terrorist bombing, and riots or active shooter) |
| Sena et al (2021) [4] | United States | Pre-post | 18 | Emergency medicine residents | 2 hours | 1 (explosion or mass shooting with fire) |
| Farhat et al (2022) [29] | Qatar | Pre-post | 12 | Physicians, nurses, and paramedics | 3 hours | 4 (biological, chemical, radiological, and nuclear CBRNE[d]) |
| Alakrawi et al (2024) [30] | Saudi Arabia | Quasi-experimental | 45 | Senior paramedic students | Not reported | 1 (multivehicle accident, 15 patients) |
| Chumvanichaya et al (2025) [31] | Thailand | RCT[e] | 83 | Paramedic students | 40 minutes | 3 (10 simulated victims: SIEVE[f], SORT[g], and START[h] protocols) |
| Cuartas-Alvarez et al (2026) [32] | Spain | Quasi-experimental | 27 | Primary care professionals (doctors and nurses) | 2 hours | 1 (MCI scenario: collapse with multiple victims; PHCT[i]-led prehospital response with ECC[j] or resource coordination) |
| Tier 3: systems integration testing (n=4) | | | | | | |
| Anan et al (2016) [33] | Japan | Quasi-experimental | 230 | Physicians, firefighters, police officers, and emergency responders | 60 - 85 minutes per scenario | 3 (fire → chemical recognition → decontamination → triage) |
| Cicero et al (2019) [34] | United States | Descriptive | 27 | EMS, nurses, physicians, hospital administrators, and emergency managers | Not reported | 1 (school bus rollover, pediatric MCI) |
| Skryabina et al (2020) [35] | United Kingdom | Mixed methods | 125 | Prehospital, hospital staff, and emergency planners | Not reported | 1 (suicide bombing + marauding terrorist firearm attack) |



XSL•FO
RenderX

| Study | Country | Design | Participants, n | Population | Duration | Scenarios |
|---|---|---|---|---|---|---|
| Zimmerman et al (2026) [36] | Sweden | Mixed methods | 16 | Physicians, residents, specialists across emergency medicine, anesthesiology, internal medicine, surgery, psychiatry, radiology, and ophthalmology | Not reported | Not reported |

[a]EMS: emergency medical services.

[b]The participant number is not clearly reported (NR); there were 253 patient cases available for triage across the scenario sets.

[c]MCI: mass casualty incident.

[d]CBRNE: Chemical, Biological, Radiological, Nuclear, Explosive.

[e]RCT: randomized controlled trial.

[f]SIEVE: Sieve (UK Major Incident Primary Triage).

[g]SORT: Secondary Triage (UK Major Incident).

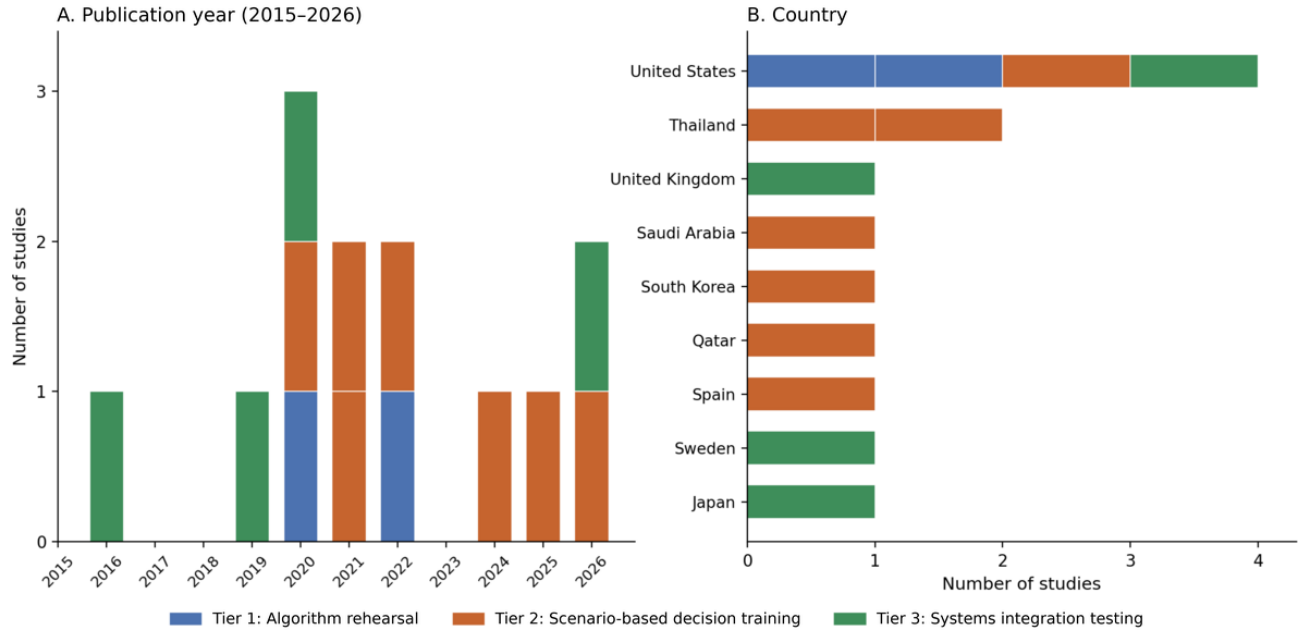
[h]START: Simple Triage and Rapid Treatment.

[i]PHCT: primary health care team.

[j]ECC: emergency coordination center.

The study populations were diverse, comprising EMS personnel, paramedics, paramedic students, physicians, emergency medicine residents, pediatric emergency specialists, nurses, and public health officers. Additionally, 3 studies specifically included nonclinical stakeholders: hospital administrators and emergency managers [34], firefighters and police officers [33], and emergency planners [35]. Table 2 provides a summary of the main study characteristics, while Figure 2A and B illustrate the distribution of the studies over time and across different regions.

**Figure 2.** Temporal and geographic distribution of included studies (N=13). Panel A shows the count of included studies by publication year from 2015 to 2026, while panel B illustrates the geographic distribution across countries. In both panels, the bar segments are color-coded to represent the Tabletop Exercise Design Spectrum tiers: tier 1 for algorithm rehearsal, tier 2 for scenario-based decision training, and tier 3 for systems integration testing.



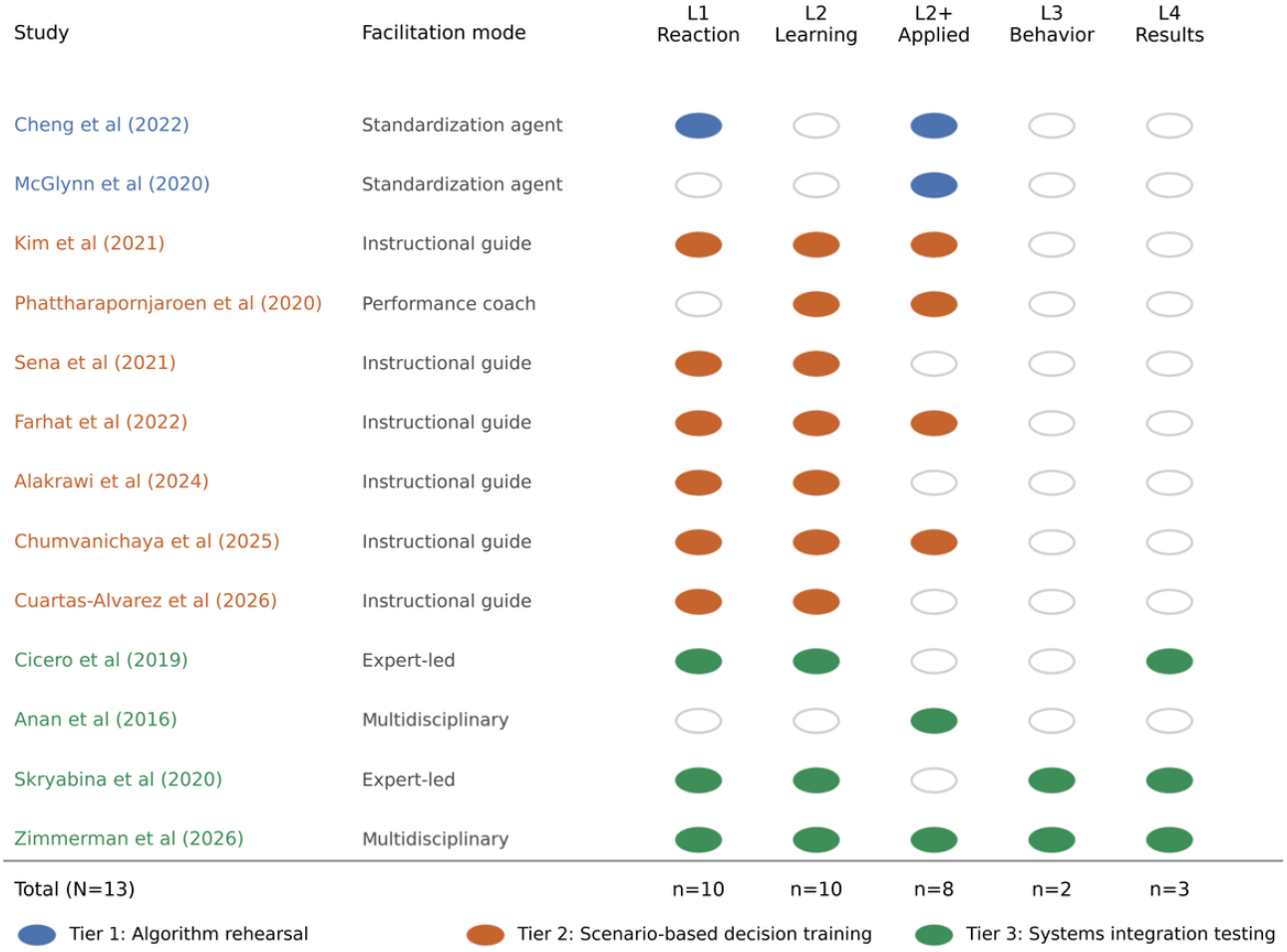

## Organization of the Synthesis

Figure 3 offers a visual overview that clarifies the relationship between the classification systems and how the results are organized. In this review, “tiers” specifically refer to the TTX Design Spectrum, which relates to exercise purpose and scope, while “levels” exclusively denote Kirkpatrick outcome levels, indicating evaluation depth. Facilitation modes describe how facilitators engaged participants. In the discussion, we use descriptive names such as algorithm rehearsal, scenario-based decision training, and systems integration testing to refer to tiers, thereby keeping the 2 numbering systems distinct.



XSL•FO
RenderX

**Figure 3.** Study-level matrix of tabletop exercise (TTX) tier, primary facilitation mode, and Kirkpatrick levels assessed (N=13). Each row represents a study included, color-coded by TTX Design Spectrum tier, highlighting its main facilitation mode and the Kirkpatrick outcome levels evaluated (level 1, Reaction; level 2, Learning; level 2+, Applied Learning; level 3, Behavior; and level 4, Results). Filled circles denote assessed levels; column totals match those in Table 3. Each study is assigned to 1 tier and 1 primary facilitation mode but can contribute to multiple outcome levels [4,25-36].




## The TTX Design Spectrum: 3 Tiers of Exercise Purpose

### *Overview*

Through iterative comparison across studies, we grouped the 13 included studies into 3 initial tiers based on each TTX's primary goal and scope. This grouping was consistent across 3 key dimensions: learning objectives, participant makeup, and outcome focus. Tier 1 involved proficiency with single-skill triage algorithms among relatively uniform prehospital groups, tier 2 covered integrated decision-making involving multiple competencies within multidisciplinary clinical teams, and tier 3 included nonclinical stakeholders, focusing on system readiness, organizational coordination, and policy changes. These tiers are preliminary patterns derived inductively and should be validated with larger samples.

### *Tier 1: Algorithm Rehearsal*

Two studies examined how individual responders interacted with a triage protocol by using TTXs to practice and assess specific procedures in controlled settings [25,26]. These exercises evaluated triage accuracy, used standardized scoring, and tested performance within short time frames. McGlynn et al [26] used patient groups of 10, 100, and 1000 casualties, while Cheng et al [25] compared 6 pediatric triage algorithms. Participants were mainly prehospital professionals trained in the relevant protocols. This phase focused on measuring individual accuracy and the correct application of algorithms rather than on team coordination or broader decision-making contexts.

### *Tier 2: Scenario-Based Decision Training*

Seven studies focused on interaction within interdisciplinary teams, using TTXs to enhance multicompetency decision-making under realistic operational constraints [4,27-32]. These exercises went beyond triage to involve incident command, team coordination, resource management, CBRNE (Chemical, Biological, Radiological, Nuclear, Explosive) response, leadership, and ethical reasoning. Duration varied from 25 minutes to 2 days, with 1 - 4 scenarios per session. Scenarios were tailored to specific contexts and included incidents such as a hydrofluoric acid chemical plant explosion [27]; fire, terrorism, and active shooter events [28]; mass shootings and fires [4]; 4 CBRNE situations [29]; a multivehicle accident [30]; triage across multiple disaster protocols [31]; and a building-collapse MCI scenario [32]. Participants were typically multidisciplinary, comprising prehospital personnel, emergency department staff, primary care providers, and sometimes public health officials. This tier prioritized integrated team decision-making under time constraints and uncertain information, rather than isolated skill assessment.



XSL•FO
RenderX

### *Tier 3: Systems Integration Testing*

Four studies examined interorganizational and agency interactions through TTXs to evaluate coordination, collaboration, and policy response strategies [33-36]. These exercises engaged multiple agencies beyond clinical staff, such as hospital leaders, emergency services, firefighters, police, and other government entities. The scenarios included pediatric MCI management after a school bus crash [34], a CBRNE response with decontamination and triage [33], a combined suicide bombing and terrorist attack involving firearms [35], and postgraduate disaster medicine training focused on leadership and system coordination [36]. Unlike tiers 1 and 2, this level prioritized organizational preparedness, clarifying agency roles, formal after-action reviews, policy documentation, memoranda of understanding, and ongoing postexercise system improvements over immediate individual training.

## Facilitation Approaches Across Tiers

Across the 13 studies, 5 recurring facilitation modes were identified, each associated with particular tiers and learning goals. These modes reflected differences in the facilitator's main role, such as guiding participants through a scenario, providing feedback on performance, ensuring consistent assessment, or supporting postexercise review. The following subsections describe the characteristics of each mode and its use across the 3 tiers.

### *Instructional Guide (Tier 2 Predominant)*

Six tier 2 studies incorporated active instructional guidance in exercise scenarios [4,27,29-32]. Facilitators managed the scenario flow by providing timed information, posing clarifying questions, and offering prompts when decision-making slowed. They also adjusted scenario elements such as casualty numbers, resource availability, and time constraints to enhance learning. After each scenario, structured debriefings took place, often led by facilitators who asked questions to reinforce essential learning points.

### *Performance Coach (Tier 2)*

The study by Phattharapornjaroen et al [28] involved placing 2 supervisory personnel in each exercise group, explicitly tasked with observing and giving iterative feedback on team leadership behaviors and decision-making quality, using a combination of live observation and structured feedback.

### *Standardization Agent (Tier 1 Predominant)*

McGlynn et al [26] and Cheng et al [25] emphasized measurement consistency and the reliability of assessments by using standardized facilitation protocols [25,26]. Facilitators underwent rater calibration training, used reference sheets and standardized scoring guides, and followed scripted scenario presentations to ensure interrater reliability.

### *Expert-Led Facilitation (Tier 3)*

Cicero et al [34] and Skryabina et al [35] involved senior clinicians and emergency preparedness specialists in conducting exercises and guiding after-action reports. Their role therefore extended beyond scenario delivery to postexercise review, consistent with tier 3's focus on organizational preparedness and interagency coordination.

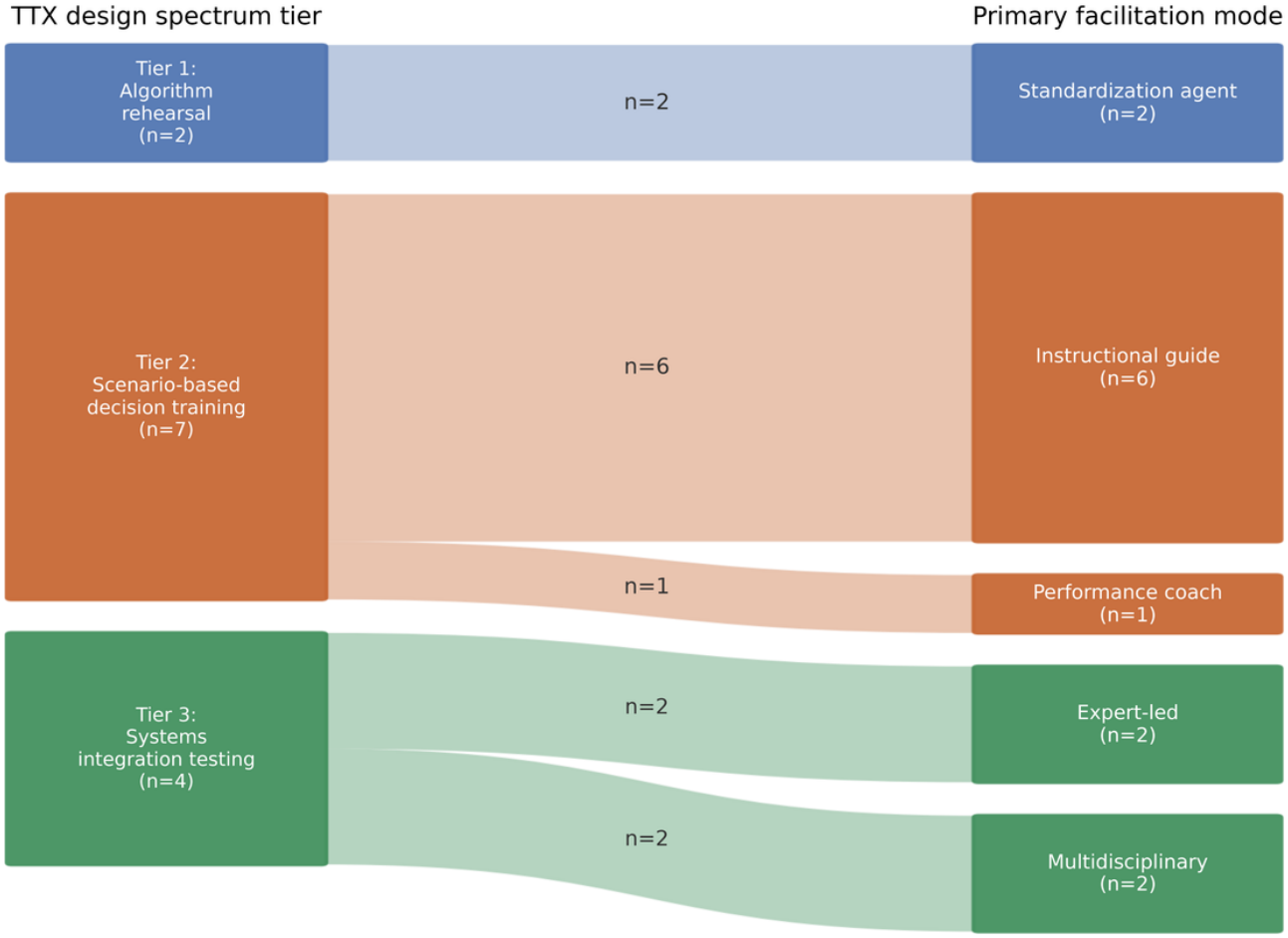
### *Multidisciplinary Facilitation (Tier 3)*

Anan et al [33] and Zimmerman et al [36] employed multidisciplinary facilitation teams from various agencies to guarantee scenario realism and credibility within each participant's operational setting. In both modes, tier 3 facilitation emphasized interagency collaboration and systems-level learning over individual skill evaluation. Figure 4 shows the alignment between TTX tiers and facilitation modes.



XSL•FO
RenderX

**Figure 4.** TTX tier–facilitation mode alignment (N=13). Alluvial diagram illustrating the connection between TTX Design Spectrum tiers (on the left) and primary facilitation modes (on the right). Blue represents tier 1 (algorithm rehearsal), brown represents tier 2 (scenario-based decision training), and green represents tier 3 (systems integration testing). The lighter-shaded flows retain the color of their source tier, and flow width corresponds to the number of studies. Each study is linked to 1 specific tier and 1 primary facilitation mode. TTX: tabletop exercise.




## Evaluation Strategies and Kirkpatrick Mapping

### *Data Collection Instruments and Timing*

The 13 studies used diverse data collection methods. Tools for Kirkpatrick levels 1 - 2 included knowledge tests (used in 8 studies), satisfaction ratings (7 studies), confidence assessments (5 studies), and self-reports on knowledge, competency, or preparedness (also 5 studies). Reaction measures extended to perceived usefulness, usability, and relevance ratings. Multimedia Appendix 4 details the specific instruments mapped to each outcome level [4,25-36]. For level 2+ (Applied Learning), assessment methods included scoring triage accuracy and speed against reference standards [25,26,31], structured observation to evaluate team behaviors, task performance, decision quality during exercises [27-29,36], and after-action reports assessed across sequential scenarios [33].

Level 3 (Behavior) evidence includes documented use of exercise-based skills in real operational contexts. Two studies reported this level of evidence. Skryabina et al [35] observed that participants used exercise-based knowledge and decision-making during the Manchester Arena bombing response. Zimmerman et al [36] described postprogram role application, with participants taking on roles such as departmental preparedness coordinator and simulation instructor-candidate, suggesting early signs of transferring practice beyond the exercise environment.

Level 4 data pertain to organizational change. Cicero et al [34] developed hospital preparedness metrics based on a 2-year postexercise follow-up. Skryabina et al [35] observed improvements in casualty distribution across the system and enhanced operational efficiency during the actual MCI response. Zimmerman et al [36] noted early organizational impacts, such as establishing a structured simulation platform, conducting additional course iterations, holding refresher meetings, engaging in alumni follow-up activities, and continuously refining contingency plans. Most studies used pre- and postassessment timings, with some also including during-exercise measurements such as triage accuracy and incident command compliance. Additionally, 1 study incorporated a 1-week delayed posttest to assess knowledge retention [31].

### *Kirkpatrick-Tier Mapping*

Across the 13 studies, the distribution of Kirkpatrick levels was as follows: level 1 (Reaction) n=10, level 2 (Learning) n=10, level 2+ (Applied Learning) n=8, level 3 (Behavior) n=2, and level 4 (Results) n=3. The level 2+ category indicates a measurement pattern that the standard 4-level model does not clearly depict. The study-level distribution of Kirkpatrick



XSL•FO
RenderX

outcome assessments across tiers is shown in Figure 3, with detailed outcome mapping by study available in Multimedia Appendix 4. Evaluation depth varied between tiers (Table 3). Outcomes for levels 3 and 4 were assessed solely in systems integration (tier 3) studies [34-36], aligning with this tier's emphasis on systems-level readiness and real-world validation opportunities.

**Table .** Kirkpatrick outcome assessment by Tabletop Exercise Design Spectrum tier (number of studies assessing each level).[a]

| Tier | Studies | Level 1: Reaction | Level 2: Learning | Level 2+ (Applied Learning) | Level 3: Behavior | Level 4: Results |
|---|---|---|---|---|---|---|
| 1: algorithm rehearsal | 2 | 1 | 0 | 2 | 0 | 0 |
| 2: scenario-based decision training | 7 | 6 | 7 | 4 | 0 | 0 |
| 3: systems integration testing | 4 | 3 | 3 | 2 | 2 | 3 |
| Total (N=13) | 13 | 10 | 10 | 8 | 2 | 3 |

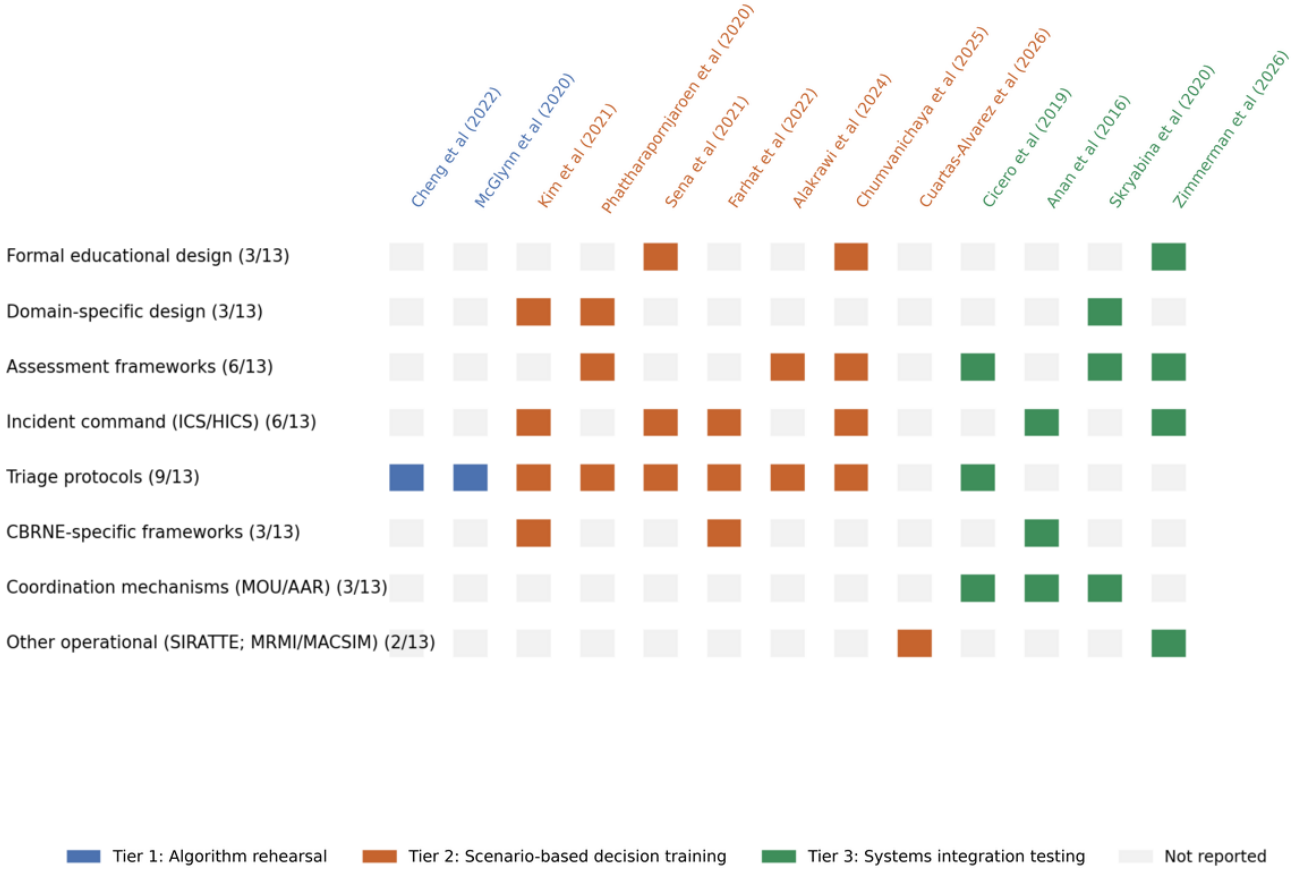
[a]Cell values represent the number of studies evaluating each level; a single study can contribute to multiple levels. Level 2+ (Applied Learning) indicates structured performance assessments conducted within the exercise environment (Multimedia Appendix 2).

## Frameworks Guiding Exercise Design and Evaluation

### *Overview*

Figure 5 provides a summary of how training design, assessment, and operational frameworks are reported across the 13 included studies. Operational frameworks were documented most consistently, while specific educational design and assessment frameworks appeared less frequently.

**Figure 5.** Framework reporting landscape across included studies (N=13). Matrix chart displaying framework reporting across studies, categorized into 8 groups: formal educational design, domain-specific design, assessment frameworks, and 5 operational areas: incident command (ICS/HICS), triage protocols, CBRNE-specific frameworks, coordination mechanisms (such as memoranda of understanding and after-action reports), and other operational frameworks (such as SIRATTE; MRMI/MACSIM). Columns represent individual studies, color-coded by TTX Design Spectrum tier; filled cells denote reported frameworks, while light gray cells show frameworks not reported [4,25-36]. AAR: after-action report; CBRNE: Chemical, Biological, Radiological, Nuclear, Explosive; HICS: Hospital Incident Command System; ICS: Incident Command System; MACSIM: Mass Casualty Simulation System; MOU: Memorandum of Understanding; MRMI: Medical Response to Major Incidents; SIRATTE: Security, Information, Roles, Areas, Triage, Treatment, and Evacuation.






XSL•FO
RenderX

### *Training Design Frameworks*

Out of 13 studies, only 3 explicitly described formal educational training frameworks. Chumvanichaya et al [31] used backward design, focusing the exercise on previously identified preparedness gaps. Sena et al [4] applied Kolb's [37] experiential learning model and adult learning principles to shape scenario development and debriefings. Zimmerman et al [36] integrated disaster medicine education and competency frameworks into postgraduate training. Seven studies did not specify any formal educational design framework [25,26,29,30,32-34]. Meanwhile, 3 studies used domain-specific frameworks instead of broad educational models: Kim et al [27] organized CBRNE response phases around the Chain of Chemical Survival, Phattharapornjaroen et al [28] applied the 3LC (3-Level Collaboration) framework with Major Incident Medical Management and Support or Medical Response to Major Incidents triage protocols, and Skryabina et al [35] based their approach on the Emergency Preparedness Cycle and Emergo Train System. Overall, most research relied more on disaster response or operational frameworks than on explicit educational design theories.

### *Assessment Frameworks*

Assessment frameworks were described more clearly than training design frameworks. Phattharapornjaroen et al [28] used the CSCATTT (Command, Safety, Communication, Assessment, Triage, Treatment, Transportation) framework to organize and assess incident command performance. Cicero et al [34] used the Hospital Pediatric Disaster Preparedness Checklist and after-action reports to evaluate organizational readiness. Skryabina et al [35] integrated Kirkpatrick levels with the Promoting Action on Research Implementation in Health Services framework to analyze both training outcomes and implementation processes. Other studies used specific structured tools for assessments: Chumvanichaya et al [31] applied the Attention, Relevance, Confidence, Satisfaction Motivation Model to measure participant engagement, Alakrawi et al [30] adapted the Competency Level Use in Triage Scale to evaluate triage competency, and Zimmerman et al [36] used a 29-item questionnaire along with CSCATTT-anchored instructor scores to assess applied performance during simulations.

### *Operational Frameworks*

Operational frameworks were reported more consistently than educational design frameworks. Six studies mentioned the Incident Command System or Hospital Incident Command System [4,27,29,31,33,36]. Triage protocols, such as Simple Triage and Rapid Treatment; Sort, Assess, Lifesaving Interventions, Treatment/Transport; Sieve (UK Major Incident Primary Triage); Secondary Triage (UK Major Incident); JumpSTART; CareFlight; Pediatric Triage Tape; and Sacco triage, were the most common operational frameworks used across different tiers [4,25-31,34]. Three studies used CBRNE-specific response frameworks: the Chain of Chemical Survival [27], CBRNE workshop competencies [29], and the MCLS-CBRNE course framework [33]. Tier 3 studies more frequently included coordination mechanisms such as memoranda of understanding, centralized dispatch protocols, interagency responsibilities, and after-action reports [33-35]. Cuartas-Alvarez et al [32] used SIRATTE, a prehospital MCI framework that covers Security, Information, Roles, Areas, Triage, Treatment, and Evacuation, in the MassCas tabletop game. Zimmerman et al [36] incorporated several operational and collaborative frameworks, such as Incident Command System, CSCATTT, 3LC, Medical Response to Major Incidents, and Mass Casualty Simulation System, into its postgraduate disaster medicine program.

## Discussion

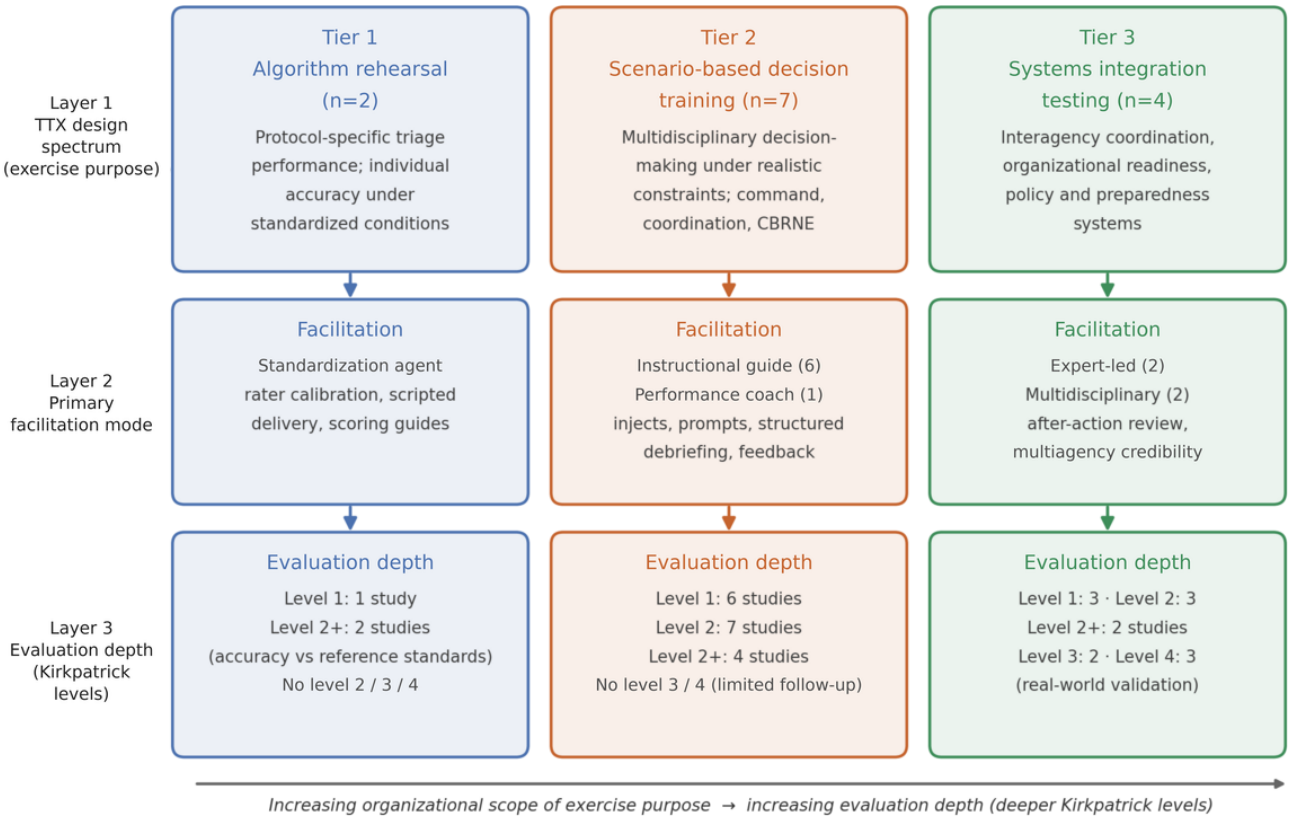
### Principal Findings

The main conclusion of this scoping review is that TTXs in prehospital MCI preparedness should be viewed as a group of related exercise types, each serving different educational and operational goals, rather than a single, uniform intervention. These purposes are categorized into 3 tiers: algorithm rehearsal, scenario-based decision training, and systems integration testing. Outcomes are mapped using the Kirkpatrick framework, with level 2+ indicating structured performance during the exercise that lies conceptually between knowledge acquisition and real-world application. These tiers represent increasing levels of interaction: between individual responders and protocols, among interdisciplinary team members, and across organizations and agencies. Facilitation modes involve intentionally designed interactions between facilitators and participants. While diverse exercise designs are common and often suitable, the key message is that the purpose, facilitation, and evaluation of the exercise must be explicitly aligned. Without a shared typology of purpose, evidence from exercises with different goals cannot be meaningfully compared. As shown in Figure 6, the literature indicates a pattern where exercise purpose, facilitation mode, and evaluation depth progress together.



XSL•FO
RenderX

**Figure 6.** Integrated overview of TTX purpose, facilitation, and evaluation depth across included studies (N=13). The figure outlines the included studies across 3 interconnected layers: (1) the TTX Design Spectrum tier, (2) the main facilitation modes associated with each tier, and (3) the evaluation depth categorized by Kirkpatrick levels, including level 2+ (Applied Learning). CBRNE: Chemical, Biological, Radiological, Nuclear, Explosive; TTX: tabletop exercise.




The tier typology should be seen as an initial guideline rather than a confirmed classification. Nonetheless, it clarifies why studies labeled as TTXs often appeared quite different in practice: algorithm-rehearsal exercises focused on standardized practice of specific skills [25,26], scenario-based decision training involved integrated clinical and command decisions in discussion settings [4,27-32], and systems-integration exercises targeted organizational coordination and system readiness [33-36]. Exercises aimed at rehearsing an algorithm, enhancing team decision-making, or testing system coordination do not share the same facilitation approach or outcome metrics.

The Design Spectrum should be aligned with existing exercise doctrine. The HSEEP (Homeland Security Exercise and Evaluation Program) connects exercise goals to structured assessments using capability-based exercise evaluation guides and standardized after-action reports [38]. However, HSEEP treats TTXs as a single, uniform discussion-based type. The current typology offers a more detailed classification within that category, identifying different types of TTXs that require varied facilitation and evaluation approaches. This typology is designed to complement HSEEP objectives and not replace them.

This has a practical implication for anyone designing or assessing a TTX: the evaluation approach should be aligned with the exercise's goal, not judged by a universal standard of assessment thoroughness. A focused protocol rehearsal may rightly prioritize structured performance metrics such as decision accuracy and interrater reliability, without necessarily claiming to measure organizational impact [25,26]. Meanwhile, a decision-training exercise might aim to improve knowledge, confidence, communication, and role clarity but should ideally incorporate follow-up methods if it seeks to influence future practice [4,29,32]. A systems-centered exercise should analyze after-action processes, coordination structures, and readiness outputs, as these are its core focus [34-36]. The key point is not whether every TTX achieves the highest Kirkpatrick level but whether the chosen level is appropriate for the capability being tested.

## Separating Applied Learning From Behavioral Transfer

The level 2+ subcategory is valuable because it differentiates between applied performance during the exercise and behavioral transfer outside of it. Multiple studies have assessed applied performance during the tabletop, providing evidence of learning in a simulated environment [25-29,31,33,36]. While this evidence is educationally significant, it does not necessarily reflect practice change in real settings. Making this distinction helps prevent overinterpreting simulation results as proof of real-world transfer. This differentiation is not new; models such as the revised Kirkpatrick framework, Miller's framework, and translational science models of simulation-based education outcomes all recognize it [22-24]. However, its systematic application to prehospital MCI tabletop evidence has not been reported before. Broader reviews have generally shown clustering at the Reaction and Learning levels without separating structured in-exercise performance from actual operational transfer [7,8].



XSL•FO
RenderX

This distinction also assists in understanding the limited occurrence of specific behaviors and outcome results. Their absence should not be automatically interpreted as failure. In prehospital disaster education, actual events are rare, unpredictable, and difficult to tie to a single training session [39], making it challenging to observe behavioral transfer even when learning has occurred. Additionally, follow-up methods in scenario-based decision-training studies were often limited [4,29,30,32], with only a single delayed posttest after 1 week [31], which hampers the ability to determine whether these exercises impacted later operational reasoning. This does not imply that such exercises are ineffective but rather that they lacked sufficient follow-up relative to their expected outcomes, highlighting a gap in evaluation data rather than a flaw in design.

### Timing, Transfer, and Facilitation

The study by Skryabina et al [35] provides rare naturalistic evidence that exercise learning can be applied in real responses when training is close to deployment. It found that participants reported using knowledge from the TTX during the Manchester Arena incident, which occurred soon afterward. Although this single uncontrolled observation cannot establish causality, it aligns with the literature on distributed practice, which shows that the spacing and timing of learning opportunities influence retention and transfer [40], and with the advice by Cicero et al [34] to exercise every 6 months [34]. While single-exposure TTXs may support learning, lasting behavior change likely depends more on repetition and deliberate timing than on a one-time exposure. This remains an open question for future research rather than a definitive conclusion of this review.

Facilitation findings consistently indicate that it is a strategic design choice rather than a neutral delivery method, and that it varies with the exercise's purpose (Figure 4). More directive, standardized facilitation is better suited for narrowly focused rehearsal activities, while expert-led, coaching, or multidisciplinary facilitation is typically used in broader scenario-based and system-level exercises [4,27,34,36]. There is no single facilitation style that is universally superior; instead, the facilitator's role should be aligned with the exercise's goals, whether to standardize performance, encourage collaborative reasoning, identify coordination gaps, or foster organizational reflection.

### Frameworks and Implementation

Operational frameworks were described more consistently than educational or assessment frameworks (Figure 5). Only a few studies explicitly detailed formal educational frameworks such as experiential learning, backward design, or competency-based disaster medicine education [4,31,36]. While the field is grounded in operational principles, its pedagogical aspects are less well articulated. Operational doctrine can effectively shape scenarios, but inconsistent reporting of educational and assessment approaches hampers reproducibility and obscures which design choices actually enhance learning or transfer. Notably, the combination of the Kirkpatrick model with the Promoting Action on Research Implementation in Health Services implementation framework by Skryabina et al [35] stands out, as it connects training outcomes with implementation processes rather than viewing evaluation as an isolated end point [35,41,42]. This approach demonstrates how exercise evaluation can more systematically contribute to system-level preparedness improvements.

These findings generally align with earlier review literature but offer a more specific interpretive perspective. Frégeau et al [7] observed outcomes primarily at lower Kirkpatrick levels in medical emergency contexts, while Emaliyawati et al [8] noted that tabletop disaster exercises enhance knowledge, confidence, preparedness, and performance. This review differs by focusing on the prehospital MCI interface, introducing the level 2+ category, and suggesting that the main challenge is purpose-aligned evaluation rather than measurement deficiencies. This distinction is practically valuable, helping educators and planners determine the kind of evidence they can reasonably expect from various exercises.

The clearest takeaway from practice is that TTXs should be designed starting from the desired capability. If the aim is to rehearse algorithms, then structured in-exercise performance measures might suffice as end points. For scenario-based decision training, evaluation should go beyond participant satisfaction and include follow-up methods such as delayed case-based testing, repeated scenarios, structured interviews, or local practice audits. If the focus is on systems integration, exercises should connect to after-action reviews, policy updates, coordination metrics, or preparedness planning [34-36]. This review advocates for a proportional evaluation approach: not every exercise needs maximum evaluation depth, but each should have an evaluation strategy aligned with its purpose. Multimedia Appendix 5 summarizes these pairings, listing typical objectives, facilitation modes, evaluation targets, example instruments from the studies, and suggested data to record [25-28,31,33-36].

### Future Research

Future research should focus on refining this purpose alignment approach instead of merely questioning whether TTXs "work" in general. One key area is to develop practical methods for measuring transfer following scenario-based decision-training exercises without relying on the occurrence of a real MCI: options include repeating scenarios, using follow-up decision vignettes, conducting interval audits of triage or command processes, and linking short-format exercises as possible next steps. Another important area is to investigate whether repeated or ongoing exposure to related TTXs enhances retention and transfer more than isolated sessions. This is especially relevant since the current review does not establish a dose-response relationship; however, the training-transfer literature suggests that this is plausible because transfer relies on reinforcement, opportunities for application, and workplace conditions [43,44].

A third priority involves stakeholder-based implementation research, engaging EMS leaders, emergency preparedness officers, public health partners, hospital interface stakeholders, and simulation faculty in designing exercise objectives and follow-up plans. This ensures that evaluations align with real operational needs. It should also include a more structured, comparable collection of evaluation data, such as observation ratings, accuracy scores, and after-action findings, using existing templates such as HSEEP's Exercise Evaluation Guides and



XSL•FO
RenderX

standardized reporting [38]. A fourth priority is expanding scenarios beyond specific, bounded incidents. Predictable hazards such as storms, wildfires, floods, or hurricanes could be especially valuable, as threat windows can sometimes be forecasted. Advances in satellite hazard assessments, artificial intelligence–enabled risk estimation, and forecasting may help identify these periods and enable more realistic timing of training relative to operational demand [39,45]. These designs could also examine whether training closer to deployment improves transfer.

### Limitations

Several limitations should inform how these findings are interpreted. First, the evidence base was limited in size and diverse in nature, so the proposed tiers and alignment patterns should be viewed as preliminary, hypothesis-generating concepts rather than confirmed frameworks. Specifically, since the tier definitions partly depended on participant composition and outcome scope, the link between tier and evaluation depth is partly inherent and should not be seen as an independent empirical correlation. Second, this was a scoping review without formal critical appraisal; thus, the findings depict the state of the literature rather than comparing the effectiveness of different TTX designs. Most studies included used pre-post, quasi-experimental, descriptive, or mixed methods approaches, which restrict causal conclusions.

Third, the study reporting was inconsistent. The definitions of outcomes, facilitation descriptions, and framework reporting varied, leading to interpretive uncertainty when aligning studies across tiers and Kirkpatrick levels. The level 2+ subcategory enhances conceptual clarity, but its assignment still relied on reviewer judgment. Fourth, the literature was mainly from high-income or upper-middle-income settings and focused on human-caused or bounded hazard scenarios, restricting its applicability to resource-limited environments and prolonged natural disasters (Figure 2). Additionally, limiting the review to English-language sources might have excluded important evidence.

Two of the review's more speculative points rely on limited evidence. The idea that temporal proximity between training and real events might facilitate transfer primarily derives from a single naturalistic case [35]. While interesting, it remains unconfirmed. Similarly, the notion that short, time-limited tabletop formats may lack the functional fidelity, according to the definition by Hamstra et al [46], to simulate long-term, evolving disasters that disrupt infrastructure is speculative. The existing literature primarily involves bounded incidents and does not directly test this idea. Both points are presented as suggestions for future research rather than confirmed explanations.

### Conclusions

TTXs for prehospital MCI preparedness should not be seen as a single, uniform approach. Their success mainly depends on how clearly the designers define the expected level of interaction. Whether it is an individual following protocols, a multidisciplinary team working together, or organizations coordinating efforts, facilitation and evaluation should be tailored accordingly. By concentrating on the prehospital MCI literature and differentiating between applied learning during exercises (level 2+) and actual behavioral transfer, while considering variations through the lens of design–evaluation alignment, this review offers a targeted synthesis of the prehospital MCI interface that complements earlier, broader tabletop reviews [7,8]. When used carefully, TTXs can remain a scalable and strategic component of prehospital disaster preparedness, with further research needed to determine when and how they promote lasting practice change and system readiness.

## Acknowledgments

The authors express gratitude to Dr Sarah Kazim, chair of emergency medicine at Dubai Health, for her support and for creating a departmental environment conducive to this work. They also appreciate Mohammed Bin Rashid University of Medicine and Health Sciences for its support and collaboration, Graduate Medical Education for assisting the residents, Shakeel Tegginmani of the Al Maktoum Medical Library for helping with the literature search, and the Institute of Learning for providing research support. Generative AI tools, such as ChatGPT and Perplexity, were used solely for editorial tasks, such as enhancing language clarity, readability, grammar, and organization, as well as brainstorming visual presentation ideas for data already extracted and analyzed by the authors. AI was not used for data generation, literature searches, study screening, data extraction, outcome classification, analysis, interpretation, or forming scientific conclusions. All AI-suggested edits were carefully reviewed, revised, and approved by the authors. The authors assume full responsibility for the accuracy, integrity, and final content of the manuscript.

## Funding

This research was supported by the Mohammed Bin Rashid University of Medicine and Health Sciences via the Dubai Health Collaborative Stimulus Research Grant for the MCIPHER project. The funding body had no role in the study design, data analysis, or manuscript writing.

## Data Availability

All data produced or examined in this study are included within this published paper and its multimedia appendices. The full data charting form can be obtained from the corresponding author upon reasonable request.



XSL•FO
RenderX

## Authors' Contributions

Conceptualization: AO, AY, NZ
Data curation: PP, AA, DN, AO
Formal analysis: AO
Investigation: PP, AA, DN
Methodology: PP, AA, DN, AO, IH, NZ
Project administration: AY
Resources: AY
Supervision: AO, AY, IH, NZ
Visualization: NZ
Writing – original draft: PP, DN, AO
Writing – review & editing: PP, AA, AY, IH, NZ

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete database-specific search strategies.
[DOCX File, 28 KB - i-jmr_v15i1e96228_app1.docx ]

Multimedia Appendix 2
Glossary, Kirkpatrick classification rules, and Tabletop Exercise Design Spectrum definitions.
[DOCX File, 39 KB - i-jmr_v15i1e96228_app2.docx ]

Multimedia Appendix 3
Metadata for unretrieved conference abstracts.
[DOCX File, 42 KB - i-jmr_v15i1e96228_app3.docx ]

Multimedia Appendix 4
Kirkpatrick outcome mapping by study and exercise tier.
[DOCX File, 38 KB - i-jmr_v15i1e96228_app4.docx ]

Multimedia Appendix 5
Practitioner summary: purpose-aligned design, facilitation, and evaluation by tabletop exercise tier.
[DOCX File, 37 KB - i-jmr_v15i1e96228_app5.docx ]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 39 KB - i-jmr_v15i1e96228_app6.docx ]

## References

1. Sendai Framework for Disaster Risk Reduction 2015-2030. United Nations Office for Disaster Risk Reduction. 2015. URL: https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030 [accessed 2026-09-15]
2. Bissmeyer H, Gallegos C, Randazzo S, Scheffler C, Sullivan Lee L, Powell L. Tabletop simulation as an innovative tool for clinical workflow testing. Nurs Adm Q 2025;49(1):E1-E7. [doi: 10.1097/NAQ.0000000000000664] [Medline: 39622034]
3. Castro Delgado R, Fernández García L, Cernuda Martínez JA, Cuartas Álvarez T, Arcos González P. Training of medical students for mass casualty incidents using table-top gamification. Disaster Med Public Health Prep 2022 Sep 21;17:e255. [doi: 10.1017/dmp.2022.206] [Medline: 36128647]
4. Sena A, Forde F, Yu C, Sule H, Masters MM. Disaster preparedness training for emergency medicine residents using a tabletop exercise. MedEdPORTAL 2021 Mar 12;17:11119. [doi: 10.15766/mep_2374-8265.11119] [Medline: 33768151]
5. Liu J, Huang Y, Li B, Gui L, Zhou L. Development and evaluation of innovative and practical table-top exercises based on a real mass-casualty incident. Disaster Med Public Health Prep 2022 May 16;17:e200. [doi: 10.1017/dmp.2022.95] [Medline: 35575292]
6. Torpan S, Orru K, Hansson S, Klaos M. Using a table-top exercise to identify communication-related vulnerability to disasters. Int J Disaster Risk Reduct 2025 Mar;119:105264. [doi: 10.1016/j.ijdrr.2025.105264]
7. Frégeau A, Vinette B, Lapierre A, et al. Tabletop simulations in medical emergencies: a scoping review. Simul Healthc 2025 Aug 1;20(4):223-228. [doi: 10.1097/SIH.0000000000000838] [Medline: 39625695]



XSL•FO
RenderX

8. Emaliyawati E, Ibrahim K, Trisyani Y, et al. Enhancing disaster preparedness through tabletop disaster exercises: a scoping review of benefits for health workers and students. Adv Med Educ Pract 2025;16:1-11. [doi: 10.2147/AMEP.S504705] [Medline: 39807178]
9. Mutiarasari D, Zulkifli A, Rivai F, Thamrin Y, Mallongi A, Miranti M. The effectiveness of table-top exercise simulation for hospital disaster preparedness training a systematic review. J Neonatal Surg 2025;14(9S):102-110. [doi: 10.52783/jns.v14.2634]
10. Khirekar J, Badge A, Bandre GR, Shahu S. Disaster preparedness in hospitals. Cureus 2023 Dec;15(12):e50073. [doi: 10.7759/cureus.50073] [Medline: 38192940]
11. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
12. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5(1):69. [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
13. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
14. Kazim M, Prashanth P, Narayanan D, et al. Tabletop exercises for prehospital preparedness during emergencies: a scoping review protocol. protocols.io. 2025. URL: https://www.protocols.io/view/tabletop-exercises-for-prehospital-preparedness-du-g4pebyvjf [accessed 2026-09-15]
15. Kazim M, Prashanth P, Narayanan D, et al. Tabletop exercises to assess prehospital preparedness: a scoping review protocol. Preprints.org. Preprint posted online on Oct 5, 2025. [doi: 10.20944/preprints202510.0457.v1]
16. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev 2021 Jan 26;10(1):39. [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
17. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 guideline statement. J Clin Epidemiol 2016 Jul;75:40-46. [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
18. Covidence systematic review software. Veritas Health Innovation. URL: https://www.covidence.org [accessed 2026-09-15]
19. Cheng A, Kessler D, Mackinnon R, et al. Reporting guidelines for health care simulation research: extensions to the CONSORT and STROBE statements. Adv Simul 2016 Jan;1(1):25. [doi: 10.1186/s41077-016-0025-y]
20. Kirkpatrick DL. Techniques for evaluating training programs. J Am Soc Train Dir 1959;13(11):3-9 [FREE Full text]
21. Kirkpatrick DL, Kirkpatrick JD. Evaluating Training Programs: The Four Levels, 3rd edition: Berrett-Koehler; 2006.
22. Kirkpatrick JD, Kirkpatrick WK. Kirkpatrick's Four Levels of Training Evaluation: ATD Press; 2016.
23. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990 Sep;65(9 Suppl):S63-S67. [doi: 10.1097/00001888-199009000-00045] [Medline: 2400509]
24. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Translational educational research: a necessity for effective health-care improvement. Chest 2012 Nov;142(5):1097-1103. [doi: 10.1378/chest.12-0148] [Medline: 23138127]
25. Cheng T, Staats K, Kaji AH, D'Arcy N, Niknam K, Donofrio-Odmann JJ. Comparison of prehospital professional accuracy, speed, and interrater reliability of six pediatric triage algorithms. J Am Coll Emerg Physicians Open 2022 Feb;3(1):e12613. [doi: 10.1002/emp2.12613] [Medline: 35059689]
26. McGlynn N, Claudius I, Kaji AH, et al. Tabletop application of SALT triage to 10, 100, and 1000 pediatric victims. Prehosp Disaster Med 2020 Apr;35(2):165-169. [doi: 10.1017/S1049023X20000163] [Medline: 32054549]
27. Kim MS, Shin H, Kim G, et al. Evaluating the Effectiveness of the Chemical-Mass Casualty Incident Response Education Module (C-MCIREM): a pilot simulation study with a before and after design. Cureus 2021 Sep;13(9):e17980. [doi: 10.7759/cureus.17980] [Medline: 34667664]
28. Phattharapornjaroen P, Glantz V, Carlström E, Dahlén Holmqvist L, Khorram-Manesh A. Alternative leadership in flexible surge capacity—the perceived impact of tabletop simulation exercises on Thai emergency physicians capability to manage a major incident. Sustainability 2020;12(15):6216. [doi: 10.3390/su12156216]
29. Farhat H, Laughton J, Joseph A, Abougalala W, Dhiab MB, Alinier G. The educational outcomes of an online pilot workshop in CBRNe emergencies. Journal of Emergency Medicine, Trauma and Acute Care 2022 Nov 22;2022(5):38. [doi: 10.5339/jemtac.2022.38]
30. Alakrawi GA, Al-Wathinani AM, Gómez-Salgado J, et al. Evaluating the efficacy of full-scale and tabletop exercises in enhancing paramedic preparedness for external disasters: a quasi-experimental study. Medicine (Baltimore) 2024 Dec 6;103(49):e40777. [doi: 10.1097/MD.0000000000040777] [Medline: 39654246]
31. Chumvanichaya K, Yuksen C, Nuanprom P, Aramvanitch K. A comparison of SIEVE, SORT, and START triage training effectiveness between immersive interactive 3D learning materials using virtual reality (VR-SSST) and traditional methods in mass casualty incidents. Int J Emerg Med 2025 Mar 13;18(1):55. [doi: 10.1186/s12245-025-00850-2] [Medline: 40082745]
32. Cuartas-Alvarez T, Garijo Gonzalo G, Valiño Otero E, et al. MassCas tabletop game as a training tool in mass casualty incidents for primary healthcare doctors and nurses: a pilot study. Prim Health Care Res Dev 2026 Mar 26;27:e39. [doi: 10.1017/S1463423626101108] [Medline: 41883338]
33. Anan H, Otomo Y, Kondo H, et al. Development of Mass-casualty Life Support-CBRNE (MCLS-CBRNE) in Japan. Prehosp Disaster Med 2016 Oct;31(5):547-550. [doi: 10.1017/S1049023X16000686] [Medline: 27531062]




XSL•FO
RenderX

34. Cicero MX, Golloshi K, Gawel M, Parker J, Auerbach M, Violano P. A tabletop school bus rollover: Connecticut-wide drills to build pediatric disaster preparedness and promote a novel hospital disaster readiness checklist. Am J Disaster Med 2019;14(2):75-87. [doi: 10.5055/ajdm.2019.0318] [Medline: 31637688]
35. Skryabina EA, Betts N, Reedy G, Riley P, Amlôt R. The role of emergency preparedness exercises in the response to a mass casualty terrorist incident: a mixed methods study. Int J Disaster Risk Reduct 2020 Jun;46:101503. [doi: 10.1016/j.ijdrr.2020.101503] [Medline: 33312855]
36. Zimmerman J, Khorram-Manesh A, Robinson Y, et al. Proactive postgraduate education in disaster medicine and preparedness for enhanced disaster management. BMC Med Educ 2026 Feb 14;26(1):336. [doi: 10.1186/s12909-026-08638-5] [Medline: 41691252]
37. Kolb DA. Experiential Learning: Experience as the Source of Learning and Development: Prentice Hall; 1984.
38. Homeland Security Exercise and Evaluation Program (HSEEP). United States Department of Homeland Security, Federal Emergency Management Agency. 2020. URL: https://www.fema.gov/emergency-managers/national-preparedness/exercises/hseep [accessed 2026-09-15]
39. Gillespie TW, Chu J, Frankenberg E, Thomas D. Assessment and prediction of natural hazards from satellite imagery. Prog Phys Geogr 2007 Oct;31(5):459-470. [doi: 10.1177/0309133307083296] [Medline: 25170186]
40. Cepeda NJ, Pashler H, Vul E, Wixted JT, Rohrer D. Distributed practice in verbal recall tasks: a review and quantitative synthesis. Psychol Bull 2006 May;132(3):354-380. [doi: 10.1037/0033-2909.132.3.354] [Medline: 16719566]
41. Kitson A, Harvey G, McCormack B. Enabling the implementation of evidence based practice: a conceptual framework. Qual Health Care 1998 Sep;7(3):149-158. [doi: 10.1136/qshc.7.3.149] [Medline: 10185141]
42. Harvey G, Kitson A. PARIHS revisited: from heuristic to integrated framework for the successful implementation of knowledge into practice. Implement Sci 2016 Mar 10;11:33. [doi: 10.1186/s13012-016-0398-2] [Medline: 27013464]
43. Baldwin TT, Ford JK. Transfer of training: a review and directions for future research. Pers Psychol 1988 Mar;41(1):63-105. [doi: 10.1111/j.1744-6570.1988.tb00632.x]
44. Grossman R, Salas E. The transfer of training: what really matters. Int J Training Development 2011 Jun;15(2):103-120. [doi: 10.1111/j.1468-2419.2011.00373.x]
45. Bari LF, Ahmed I, Ahamed R, et al. Potential use of artificial intelligence (AI) in disaster risk and emergency health management: a critical appraisal on environmental health. Environ Health Insights 2023;17:11786302231217808. [doi: 10.1177/11786302231217808] [Medline: 38089525]
46. Hamstra SJ, Brydges R, Hatala R, Zendejas B, Cook DA. Reconsidering fidelity in simulation-based training. Acad Med 2014 Mar;89(3):387-392. [doi: 10.1097/ACM.0000000000000130] [Medline: 24448038]


## Abbreviations

**CBRNE:** Chemical, Biological, Radiological, Nuclear, Explosive
**CSCATTT:** Command, Safety, Communication, Assessment, Triage, Treatment, Transportation
**EMS:** emergency medical services
**HSEEP:** Homeland Security Exercise and Evaluation Program
**MCI:** mass casualty incident
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**PRISMA-S:** PRISMA Extension for Reporting Literature Searches
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
**3LC:** 3-Level Collaboration
**TTX:** tabletop exercise


*Edited by TDA Cardoso; submitted 27.Mar.2026; peer-reviewed by C Clark, J Sands; revised version received 13.Aug.2026; accepted 14.Aug.2026; published 05.Oct.2026.*

*Please cite as:*
*Prashanth P, AlRahma A, Narayanan D, Omayer A, Yousif A, Hubloue I, Zary N*
*Design, Facilitation, and Evaluation of Tabletop Exercises for Prehospital Mass Casualty Preparedness: Scoping Review*
*Interact J Med Res 2026;15:e96228*
*URL: https://www.i-jmr.org/2026/1/e96228*
*doi:10.2196/96228*


© Paurnami Prashanth, Ali AlRahma, Dayol Narayanan, Abu Omayer, Azza Yousif, Ives Hubloue, Nabil Zary. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 5.Oct.2026. This is an open-access article



XSL•FO
RenderX

distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Application of Machine Learning Algorithms to Health Systems Planning and Public Policy Using Real-World Data: Scoping Review

Deusa Maria Arruda Rocha[1,2], MPH; Gentil Veloso Barbosa[2], PhD; Paulo Henrique De Souza Bermejo[3], PhD

[1]Department of Regulation, Evaluation, and Control (DRAC), Secretariat of Specialized Healthcare (SAES), Ministério da Saúde, Esplanada dos Ministérios, Bloco G, Edifício Anexo, Brasília, Federal District, Brazil

[2]Professional Postgraduate Program in Governance and Digital Transformation, Federal University of Tocantins, Palmas, Tocantins, Brazil

[3]Sinerji Institute for Technological and Business Research and Development, Brasília, Federal District, Brazil

**Corresponding Author:**
Deusa Maria Arruda Rocha, MPH
Department of Regulation, Evaluation, and Control (DRAC), Secretariat of Specialized Healthcare (SAES), Ministério da Saúde, Esplanada dos Ministérios, Bloco G, Edifício Anexo, Brasília, Federal District, Brazil

## Abstract

**Background:** The rapid advancement of information technology and machine learning (ML) has garnered significant interest from health care managers and policymakers. However, incorporating reliable and transparent predictions into health systems remains challenging, necessitating an investigation into gaps in digital health interventions within real-world settings characterized by resource constraints and reliability issues.

**Objective:** This study aimed to map and synthesize the available evidence on the application of ML algorithms to real-world data in health systems and identify their operational targets and clinical implications for health planning and public policy formulation.

**Methods:** Following the Joanna Briggs Institute (JBI) methodology for scoping reviews and PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines, we searched the PubMed, Embase, Scopus, and LILACS databases for empirical studies published between 2016 and 2025 that evaluated the application of ML to health management, surgical queues, and public policy using real-world data. The study protocol was retrospectively registered with the OSF.

**Results:** Of the studies identified through the database searches (2016-2025), only 6 empirical studies met all inclusion criteria, revealing that the real-world deployment of ML in health management and public policy remains at an early stage. The included models addressed micro-operational tasks, such as surgical blood transfusion forecasting, early warning system mortality prediction, postsurgical discharge planning, laboratory error detection, patient-reported outcome measure (PROM) prediction, and emergency care classification. Although high discriminatory performance was reported (area under the receiver operating characteristic curve [AUROC] 0.76-0.99), the studies were limited by single-center retrospective designs, a lack of independent external validation, and inadequate reporting of model calibration.

**Conclusions:** Current evidence on ML for macrolevel health planning and public policy formulation is limited and micro-operationally focused. Translating these algorithms into public health governance (eg, waiting list regulation in universal health systems such as the Sistema Único de Saúde [SUS]) requires moving beyond “black-box” predictions toward explainable AI (XAI), natural language processing (NLP) to analyze unstructured data, transparent model calibration, and rigorous evaluations of health equity.

***(Interact J Med Res 2026;15:e96779)*** doi:10.2196/96779

**KEYWORDS**

artificial intelligence; AI; machine learning; digital health; health management; waiting lists; elective surgical procedures; public health systems; Brazil’s unified health system; Sistema Único de Saúde; SUS

## Background

The emergence of AI as a tool for enhancing health care offers unprecedented opportunities to optimize clinical outcomes, reduce costs, and positively impact population health [1]. Among its subfields, machine learning (ML) stands out for its use of algorithmic structures and statistical modeling to identify patterns in real-world data (RWD), encompassing clinical, behavioral, and demographic variables. This facilitates the establishment of functional relationships that significantly



XSL•FO
RenderX

improve predictive accuracy [2-4]. AI-driven interventions provide robust decision support for medical professionals and create new opportunities for therapeutic monitoring, particularly in specialized care [5,6].

This technological evolution is part of a data-driven digital revolution, marking the transition from Industry 4.0 to Industry 5.0. In the health care sector, this phase—known as "Health 5.0" (H5.0)—explores technological potential to reshape service delivery, promoting personalized medicine and, most importantly, patient-centered care [2,7-10]. Interdisciplinary interest in these technologies grows alongside the global need to optimize health care spending, which accounted for approximately 11% of global gross domestic product (GDP) in 2023. This requires IT infrastructures and algorithms capable of generating clinically reliable and actionable predictions [11,12].

However, despite the exponential growth of ML research, a gap persists in both the scientific literature and government agendas regarding the effective integration of these techniques into service delivery and public policy [2,6]. This lack of standardized guidelines leaves many nations in the early stages of technological adoption [13].

A critical example is the management of waiting lists for elective surgeries, which represents a growing challenge because of population aging and the increased demand for complex procedures [3]. In this scenario, it is imperative for managers to adopt ML models to streamline workflows, ensure patient safety, and expedite care delivery. As noted by Lungu et al [14], waiting times are a central public concern and a priority for health policymakers. Yet, planning effective interventions to address waiting times remains challenging because of demographic shifts and increasing public confidence in surgical treatments.

Consequently, there is an urgent need for research addressing the intervention gap regarding the application of these tools in real-world settings. A critical limitation of the current literature is the predominance of idealized, laboratory-based simulations that fail to reflect the operational friction, missing data, and resource constraints inherent to day-to-day health system management. Demonstrating the true utility of ML in clinical practice and health governance requires evaluating models deployed within these complex real-world environments [2,15].

To address these gaps, this scoping review was guided by the following research questions:

1. How are ML algorithms currently being applied in real-world health care settings to support health systems planning and public policy formulation?
2. What are the main methodological characteristics, predictive performance measures, and clinical or operational targets of these deployed ML models?
3. What are the key gaps, methodological limitations, and barriers (eg, model calibration, external validation, and explainability) that prevent the translation of micro-operational ML models into macrolevel public health policies?

## Objective

Given the need for evidence that transcends theoretical models and addresses the complexity of health management, this scoping review aims to map and synthesize the application of ML algorithms to RWD. Specifically, we aim to identify how these tools are used in health systems planning, elective surgical flows, and public policy formulation, while characterizing the methodological landscape and identifying the limitations of the current literature.

## Methods

### Overview

This scoping review was conducted in accordance with the Joanna Briggs Institute (JBI) methodology for scoping reviews [16] and reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines [17]. The completed PRISMA-ScR checklist is provided in Checklist 1. The study focused on pragmatic, nonpharmacological health interventions incorporating ML tools and algorithms to address gaps in clinical and public health contexts.

### Protocol Registration and Open Science Transparency

The protocol for this scoping review was registered retrospectively on OSF (ucefv) [18]. Search syntaxes and documentation for both the initial search and the 2025 update were deposited and are publicly accessible via searchRxiv [19-22] and OSF repositories, ensuring full auditability and adherence to open science practices and PRISMA-ScR standards.

### Search Strategy

#### *Overview*

The search syntax and definition of the search time frame were built and expanded upon the methodological benchmark framework established by Triantafyllidis and Tsanas [2] for ML in digital health. To capture the recent decade's acceleration in the development of health management algorithms while maintaining methodological continuity, the search time frame was defined from 2016 through 2025. The search was also expanded to include LILACS to cover regional public health systems.

The systematic search was conducted in 2 distinct phases: an initial search in August 2020 and a comprehensive update in June 2025.

#### *2020 Search Phase*

The initial search strategy was developed by an experienced health librarian and structured as follows: (1) search string: "machine learning" AND "patient monitoring;" (2) period: 2016 to 2020; and (3) databases: MEDLINE (via PubMed), Embase, and Scopus. Specific details and results for this phase (n=333) are available in Table 1 and Multimedia Appendix 1.



XSL•FO
RenderX

**Table .** Quality assessment of included studies using the Effective Public Health Practice Project (EPHPP) quality assessment tool[a].

| Studies | Selection bias | Study design | Confusion factors | Blinding | Data collection methods | Withdrawals and dropouts | Global assessment |
|---|---|---|---|---|---|---|---|
| Campagner et al [23], 2020 | Moderate | Moderate | Weak | Weak | Moderate | Weak | Weak |
| Ye et al [24], 2019 | Moderate | Weak | Moderate | Weak | Moderate | Weak | Weak |
| Fontana et al [25], 2019 | Moderate | Weak | Moderate | Weak | Strong | Weak | Moderate |
| Karhade et al [26], 2018 | Weak | Weak | Weak | Weak | Strong | Weak | Weak |
| Hayn et al [27], 2017 | Weak | Moderate | Weak | Weak | Strong | Weak | Weak |
| Krämer et al [28], 2017 | Weak | Weak | Weak | Weak | Strong | Weak | Weak |

[a]This table was adapted from the study by Triantafyllidis and Tsanas [2]. Ratings are based on the EPHPP tool, where "strong" indicates no weak ratings, "moderate" indicates 1 weak rating, and "weak" indicates 2 or more weak ratings in the specific domains.

### *2025 Update Phase*

The 2025 update expanded the scope to capture the latest advancements. The search strategy, developed by a health librarian, used the following 2 main conceptual blocks:

- Block 1 (AI and ML) included the MeSH term "Machine Learning" and keywords such as "artificial intelligence," "deep learning," "neural network," and "predictive analytics."
- Block 2 (health management) focused on organizational aspects and included MeSH terms such as "Health Planning" and "Organizational Policy," combined with keywords such as "integrated care" and "health systems."

This phase was conducted in June 2025 across 4 databases: MEDLINE (via PubMed), Embase, Scopus, and LILACS. The expanded strings are available via searchRxiv [19-22] and are presented in Table 2 and Multimedia Appendix 2.



XSL•FO
RenderX

**Table .** Characteristics and predictive performance of the included studies on machine learning applications in health care management[a].

| Items | Campagner et al [23], 2020 (Italy) | Ye et al [24], 2019 (United States) | Fontana et al [25], 2019 (United States) | Karhade et al [26], 2018 (United States) | Hayn et al [27], 2017 (Austria) | Krämer et al [28], 2019 (Germany) |
|---|---|---|---|---|---|---|
| Application (objective) | Predict surgical invasiveness and recommend optimal procedures | Validate an ML[b]-based early warning system for in-hospital mortality | Predict the risk of not achieving MCID[c] in PROMs[d] after total knee arthroplasty | Predict nonroutine discharge after lumbar surgery to reduce hospital stay duration | Predict red blood cell transfusion volume for PBM[e] | Predict emergency department vs elective care urgency |
| ML algorithms | Random forest (best-performing model), ordinal regression, SVM[f], LASSO[g], KNN[h], and decision trees | Random forest (best-performing model), XGBoost[i], gradient boosting, SVM, LASSO, and KNN | Logistic LASSO, random forest, and linear SVM | ANN[j] (best-performing model), SVM, BPM[k], and decision tree; random forest used for variable selection | Random forest, multivariate logistic regression, linear models, and the Mercuriali algorithm | Random forest and KNN (for data imputation) |
| Key results (performance) | AUC[l] of 0.87 for the invasiveness score and 0.76 for procedure invasiveness | Strong discrimination (AUC 0.884); mean lead time of 40.8 hours | Fair-to-good ability to predict postsurgical MCID achievement | ANN outperformed the other models (PPV[m] 33%; sensitivity 54%) | Best-performing algorithm showed a moderate correlation (correlation coefficient=0.61) | 96% accuracy and excellent discrimination (AUROC[n]>0.99) |
| Statistical significance and evaluation metrics | Statistically significant ($P$=.04) | High AUC, indicating robust model validity | AUROC used as the primary evaluation metric | Superior AUROC, calibration, and Brier score | High statistical significance ($P$<.001) | High precision and AUROC demonstrate validity |
| Clinical implications | High physician acceptance and consistent predictive variables | Improves patient safety through early detection of clinical deterioration | Focus on MCID strengthens practical clinical applicability | Open-access tool available to support replicability and validation | Optimize PBM; applicable to various medical fields | Improves patient flow management in emergency services |

[a]This table was adapted from the study by Triantafyllidis and Tsanas [2].

[b]ML: machine learning.

[c]MCID: minimal clinically important difference.

[d]PROM: patient-reported outcome measure.

[e]PBM: patient blood management.

[f]SVM: support vector machine.

[g]LASSO: least absolute shrinkage and selection operator.

[h]KNN: k-nearest neighbors.

[i]XGBoost: extreme gradient boosting.

[j]ANN: artificial neural network.

[k]BPM: Bayes point machine.

[l]AUC: area under the curve.

[m]PPV: positive predictive value.

[n]AUROC: area under the receiver operating characteristic curve.

## Eligibility Criteria

To capture the full spectrum of relevant literature, our search strategy encompassed 4 major databases (PubMed, Embase, Scopus, and LILACS) and covered the period from 2016 through 2025. The inclusion criteria required empirical studies that applied ML algorithms specifically to health services management, care regulation, surgical queue prioritization, or public policy formulation using RWD.

The inclusion and exclusion criteria were operationalized to exclude broad, nonapplied research. Studies were excluded if they (1) evaluated ML exclusively in silico or using synthetic or benchmark datasets, without validation against real-world health system records; (2) focused exclusively on clinical diagnostic accuracy (eg, radiological image classification or genomic sequencing) without linking algorithmic outputs to operational planning, bed allocation, or administrative decision-making; (3) consisted of conceptual frameworks, reviews, editorials, or conference abstracts without full empirical implementation data; or (4) examined closed commercial software without transparent algorithmic reporting.

## Study Selection and Screening

The selection process was conducted as follows: for 2020 records, a total of 333 were identified and imported into EndNote Basic (Clarivate Analytics) [29], of which 4 duplicate records were removed. For 2025 records, a total of 611 were identified and imported into the Rayyan (Rayyan Systems Inc) web application (software as a service) [30], of which 14 duplicate records were removed.



XSL•FO
RenderX

### Consolidation and Final Selection

A total of 944 records were identified across both phases. After removing all duplicate records (n=18), 926 records underwent independent title and abstract screening. Of these, 891 were excluded. For the remaining 35 reports, full-text retrieval was successfully achieved (100%). During the eligibility assessment, 29 reports were excluded (eg, because they were reviews, lacked technical detail, or inappropriate publication date). Ultimately, 6 studies met all the criteria and were included. The complete process is illustrated in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.

### Data Extraction and Methodological Quality Assessment

Data extraction was performed using a standardized charting form inspired by the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) statement [31], focusing on key predictive performance metrics such as area under the receiver operating characteristic curve (AUROC), sensitivity, and specificity. Methodological quality was independently assessed using the Effective Public Health Practice Project (EPHPP) quality assessment tool [32,33] by a multidisciplinary team of specialists in computer science, dentistry, psychology, and statistics.

### Data Synthesis and Mapping

The data synthesis and mapping process followed a structured approach. Detailed information was extracted regarding the study objectives, methodological designs, sample sizes, and follow-up durations. Additionally, primary outcomes, reported levels of statistical significance, specific ML algorithms used, and their practical implications for clinical practice and public policy formulation were systematically mapped.

### Ethical Considerations

As this study is a scoping review of previously published literature and publicly available data, it did not involve direct interaction with human participants or the collection of identifiable private information. Consequently, formal institutional review board approval was not required. The research was conducted in accordance with the PRISMA-ScR guidelines to ensure transparency, reproducibility, and ethical scientific conduct (Checklist 1).

## Results

### Literature Search

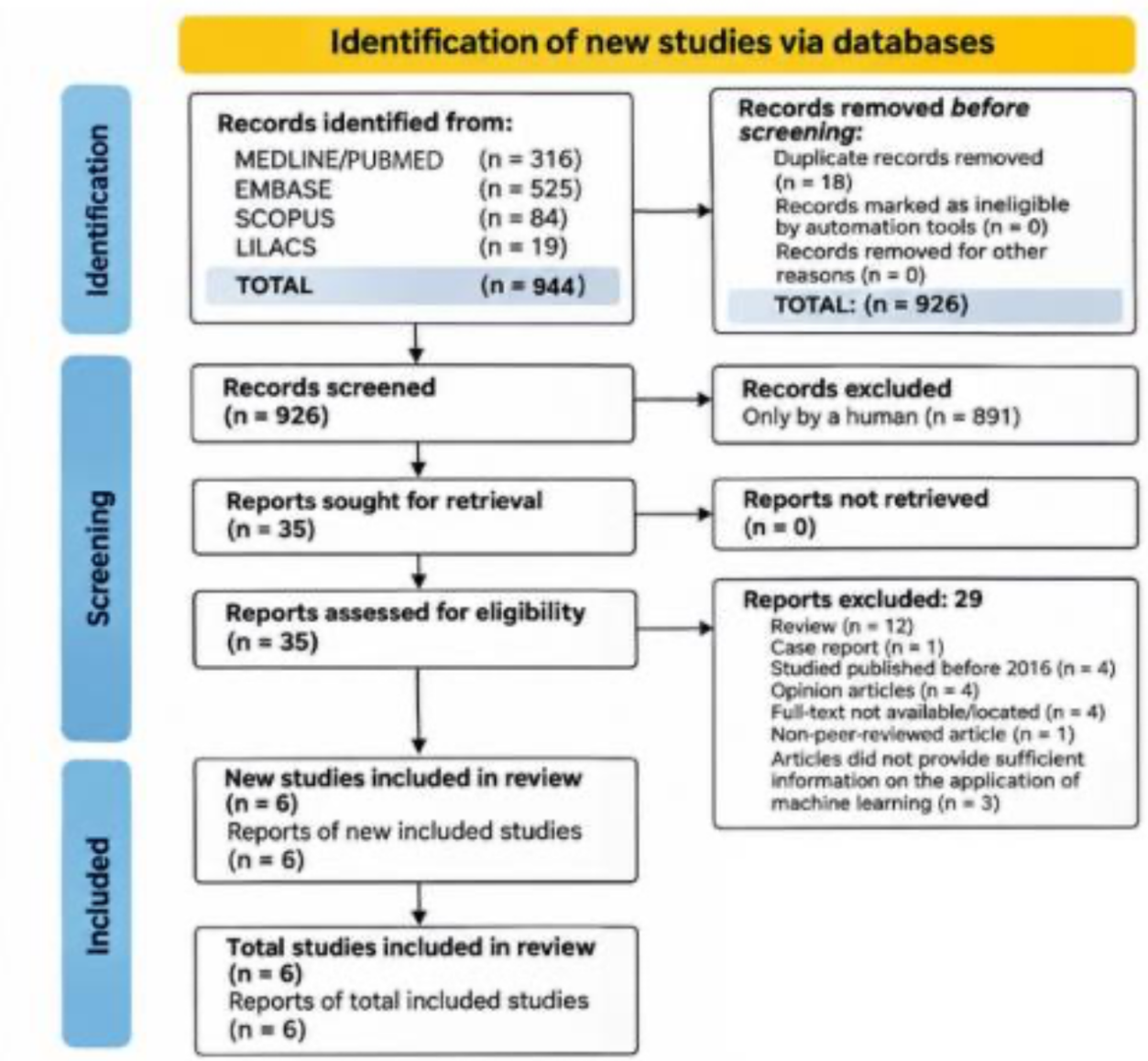
The systematic search, conducted in 2 phases (2020 and 2025), yielded a total of 944 records across the MEDLINE (via PubMed), Embase, Scopus, and LILACS databases. After removing 18 duplicate records, 926 records underwent an initial title and abstract screening. Of these, 35 articles were selected for full-text review, with a 100% retrieval success rate.

Following the eligibility assessment, 29 reports were excluded for specific reasons (eg, reviews, case reports, or lack of technical detail regarding ML implementation). Ultimately, 6 studies met all the criteria and were included in the review. These studies originated from the United States (n=3), Austria (n=1), Germany (n=1), and Italy (n=1). The complete selection process is illustrated in the PRISMA flow diagram (Figure 1).



XSL•FO
RenderX

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram for updated systematic reviews, which included searches of databases. Adapted from the study by Page et al [34], reflecting the structure of the PRISMA 2020 flow diagram adapted for this study's database search.




## Quality Assessment

Study quality was assessed using the EPHPP tool. It was observed that 83% (5/6) [23,24,26-28] of the studies presented "weak" methodological quality, while only 17% (n=1) [25] achieved a "moderate" rating. The primary weaknesses identified were selection bias and a lack of detailed information regarding participant recruitment. Regarding study design, observational cohorts prevailed (5/6, 83%) [30-34], with only 1 study characterized as a nonrandomized experimental design [23]. These findings are detailed in Table 1.

## Intervention Type and Target Population

The samples varied significantly, using sources ranging from retrospective electronic health records to administrative insurance billing data. Interventions focused on patients undergoing elective surgeries (spine, hip, and knee), emergency care, red blood cell transfusion volume management, and urgency or emergency triage classification. The average follow-up duration across the studies was approximately 5 years.

## ML Applications and Clinical Outcomes

The identified ML applications were predominantly directed toward the development of predictive models and clinical decision support systems (CDSS). As synthesized in Table 2, key application objectives encompassed risk prediction, such as modeling in-hospital mortality and nonroutine discharge following lumbar surgeries, and surgical planning, which focused on evaluating invasiveness levels and estimating required red blood cell transfusion volumes. Furthermore, technologies were implemented in triage and diagnosis, including emergency department urgency classification and predicting minimal clinically important differences (MCID) following joint arthroplasties. Algorithms used across studies included artificial neural networks (ANNs), random forests (RF), support vector machines (SVMs), and other predictive prognostic models. Although 33% of the studies reported high



XSL•FO
RenderX

statistical significance ($P$<.001), a notable lack of metric standardization was evident. Regarding performance, the models demonstrated robust results, with area under the curve (AUC) values ranging from 0.76 to 0.99, indicating excellent discriminatory performance. These clinical implications highlight the expanding role of AI tools as modalities complementary to medical judgment, fostering data-driven and evidence-based practice.

# Discussion

## Principal Findings

A critical reflection on the yield of this scoping review reveals that identifying only 6 empirical studies meeting all the strict inclusion criteria is highly informative. Rather than indicating a limitation of the search strategy, this restricted yield directly reflects an emerging evidence base in digital health. Although the theoretical literature on ML in health care is vast, true real-world implementation of ML for macrolevel health planning and public policy remains in its infancy. Most existing algorithms remain confined to laboratory simulations or isolated micro-operational tasks without integrating administrative, regulatory, or policy end points. The scarcity of published empirical deployments highlights a current structural gap between technical algorithmic development and its actual integration into public health governance, such as national waiting list regulation in universal health systems such as Brazil's Sistema Único de Saúde (SUS). Recognizing this field as an emerging area of research underscores the imperative to shift future research frameworks toward implementation science, model generalizability, and transparent calibration [6,35,36].

The empirical evidence reveals that ML use in health planning is predominantly anchored in micro-operational, clinical, and process management decision support. In elective surgical settings, models were used to optimize resource allocation and anticipate operational bottlenecks. Hayn et al [27] demonstrated that ML significantly outperformed routine practice in predicting surgical blood transfusion volumes for inventory management, while Karhade et al [26] showed that preoperative identification of nonroutine discharges mitigates financial waste and bed turnover. For long-term outcomes, Fontana et al [25] applied logistic least absolute shrinkage and selection operator (LASSO) regression (AUROC 0.60-0.89) to predict the preoperative achievement of the MCID in patient-reported outcome measures (PROMs), supporting shared decision-making. In diagnostic and systemic regulation, Campagner et al [23] applied RF models (AUROC 0.76-0.87) to administrative and laboratory data to identify sample identification errors, potentially reducing unnecessary retesting and operational costs. At the macropolicy level, Krämer et al [28] used an RF (accuracy=96%; AUROC>0.99) with data from Germany's national inpatient database (which covered 99.84% of inpatient cases) to classify emergency and elective admissions. They revealed that emergency department growth was concentrated in low-urgency cases driven by hospital reimbursement incentives (diagnosis-related groups), proving ML's utility for regulatory auditing and financial incentive design. Additionally, Ye et al [24] investigated in-hospital mortality prediction to support early warning systems for acute patient flow management.

These micro-operational ML outputs aggregate directly into 3 primary macrogovernance mechanisms: surgical queue regulation, by aggregating patient-level discharge predictions to optimize bed turnover; supply chain planning, by forecasting regional blood component demand (Hayn et al [27]); and emergency care streamlining, by stabilizing patient flow management through automated error detection and risk triage (Krämer et al [28]). To align with evidence-informed policy frameworks, ML deployment in universal health systems must address social determinants of health (SDOH) and health equity [37]. As shown by Fontana et al [25], socioeconomic variables influence predictive outcomes, posing risks that algorithms might unfairly prioritize patients with higher baseline probabilities of success over socially vulnerable individuals. Furthermore, data integrity at the source, including within systems such as e-SUS primary health care (PHC) [38], must be prioritized to avoid automation bias (Campagner et al [23]). Without continuous bias auditing, regulatory impact assessments, and robust data governance, including compliance with the General Data Protection Regulation (GDPR) [39] or Lei Geral de Proteção de Dados Pessoais (LGPD) [40], predictive algorithms risk amplifying historical disparities in access to care.

## Limitations

### *Methodological Vulnerabilities and Performance Limits*

Interpreting these findings in the context of public policy requires caution. A TRIPOD-aligned appraisal revealed high discriminatory performance (AUROC 0.76-0.99) but a pervasive omission of formal model calibration measures, such as calibration slopes, intercepts [35], or decision curve analysis. The upper bound (AUROC 0.99 in Krämer et al [28]) reflects a narrow binary triage task, whereas complex health management tasks exhibited more modest discrimination (AUROC 0.76-0.84 in Karhade et al [26]). In public health systems such as SUS, deploying miscalibrated models introduces severe risks of inappropriate queue prioritization or resource misallocation. Furthermore, 83% (5/6) of the studies exhibited overall weak methodological quality under the EPHPP tool [32,33], primarily due to high selection bias in single-center retrospective samples and unadjusted confounding. Only Campagner et al [23] achieved a moderate rating through internal validation, although still limited to a single institution. Current evidence thus represents preliminary proof-of-concept evidence of feasibility rather than evidence supporting generalizable decision-making. Scalability requires prospective multicenter validation, adherence to TRIPOD-AI [41] and PROBAST (Prediction Model Risk of Bias Assessment Tool)–AI guidelines [42], and compliance with FAIR (findable, accessible, interoperable, and reusable) data-sharing principles [43].



XSL•FO
RenderX

### *Implementation Barriers Across the Nonadoption, Abandonment, and Scale-up, Spread, and Sustainability Framework*

Analyzing implementation barriers through the nonadoption, abandonment, and scale-up, spread, and sustainability (NASSS) framework [6] shows that ML operates primarily as a decision support technology rather than an autonomous system. Regarding the technology and value proposition [44,45], the lack of external validation and calibration transparency [35] restricts technical trust. Small sample sizes relative to high feature dimensionality and single-institution coding biases further compromise transferability. Concerning adopters, the "black-box" nature [46] of complex models—such as neural networks in Karhade et al [26] or RF in Ye et al [24], Campagner et al [23], and Krämer et al [28]—fosters clinical skepticism. Multicollinearity further obscures coefficient interpretation (Fontana et al [25]), underscoring the need for explainable AI (XAI) tools [47,48] such as SHAP [49] (Shapley additive explanations) or LIME (local interpretable model-agnostic explanations), at the point of care. At the organizational level, data fragmentation across electronic health records and administrative systems hinders seamless integration. Furthermore, continuous real-time monitoring faces an "intervention paradox," as described by Ye et al [24] and Campagner et al [23], wherein successful clinical interventions avert adverse outcomes, thereby dynamically altering the model's performance metrics into statistical false positives.

### Future Directions

Bridging the current implementation gap requires 2 key future research priorities. First, adopting XAI [47,48] is critical to move beyond global discrimination metrics such as AUROC toward interpretable feature attributions, thereby ensuring ethical accountability and clinical trust. Second, integrating natural language processing (NLP) [50] is necessary to overcome the limitation of current studies, which relied exclusively on structured tabular data while ignoring rich clinical narratives. Deploying domain-adapted NLP architectures, such as ClinicalBERT, enables the automated parsing of unstructured free-text records, including referral letters, emergency intake notes, and cancellation logs in regulation systems such as SisCNRAC. Structuring unstructured text converts administrative logs into actionable intelligence, optimizing waitlist regulation, syndromic surveillance, and resource allocation across public health networks.

## Conclusions

This scoping review synthesizes the empirical baseline of ML applications deployed for health system management and public policy execution. The evidence demonstrates that real-world applications remain in an incipient phase, operating predominantly as micro-operational prediction tools for surgical queue optimization, emergency triage, postdischarge destination planning, and blood component supply management. Although these operational algorithms offer preliminary feasibility for local resource optimization, the predominance of single-center retrospective cohorts and overall weak methodological quality—as appraised via EPHPP—prevent immediate, uncritical scalability to macrolevel health governance. Bridging the gap between micro-operational predictions and systemic health policy requires transitioning to prospective multicenter validations, adopting transparent reporting guidelines (TRIPOD-AI and PROBAST-AI), enforcing strict ethical data governance, and integrating domain-specific NLP to structure real-world health records.

## Acknowledgments

The authors express their gratitude to the Secretariat of Specialized Healthcare (SAES) and the Department of Healthcare Regulation, Evaluation and Control (DRAC) for their institutional support, with special thanks to Juliana F Lujan and Débora Spalding Verdi for providing the necessary resources and support for this research. We are grateful to Departamento de Informática do Sistema Único de Saúde, Secretaria de Informação e Saúde Digital, and Ministério da Saúde for the development and support of the National High Complexity Regulation Center Information System (SisCNRAC). We also thank the Professional Doctorate Program in Governance and Digital Transformation at the Federal University of Tocantins (Universidade Federal do Tocantins), in partnership with the University of Brasília. Finally, the authors thank the Library of the Ministry of Health (Biblioteca Virtual em Saúde do Ministério da Saúde) and the National School of Public Health (Escola Nacional de Saúde Pública/Fiocruz) for their expert technical assistance with literature searching and bibliographic normalization.

## Generative AI Use

During the preparation of this work, the authors used a generative AI model, Gemini (Google) solely for technical language translation, syntactic refinement, and vocabulary optimization when translating the manuscript from Portuguese into academic English. The AI tool was not used to generate scientific concepts, synthesize data, conduct statistical analyses, or formulate research conclusions. Following translation and refinement, the authors thoroughly reviewed, edited, and validated all content to ensure scientific accuracy, and they take full responsibility for the final content of the manuscript.

## Funding

The Brazilian Ministry of Health supported this study through the professional doctorate program. The funder had no role in the study design, data collection, analysis, interpretation of data, or writing of the manuscript. The authors had full access to all the data in the study and take responsibility for data integrity and the accuracy of the data analysis.



XSL•FO
RenderX

## Authors' Contributions

DMAR contributed to the study's conceptualization, methodology, and formal analysis of the results. She led the literature screening and data extraction and drafted the original manuscript. GVB acted as the primary supervisor, contributing to the study's conceptualization and methodological oversight and providing critical revisions to the intellectual content. PHDSB contributed to the validation of the findings and the critical review of the final manuscript, offering strategic insights into governance and digital transformation policies. All authors have read and approved the final version of the manuscript and are accountable for all aspects of the work.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Initial search strategy and search strings.
[DOCX File, 114 KB - i-jmr_v15i1e96779_app1.docx ]

Multimedia Appendix 2
Search strategies deposited in the searchRxiv repository [19-22]; this appendix also contains the final MEDLINE search strategy.
[DOCX File, 52 KB - i-jmr_v15i1e96779_app2.docx ]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 109 KB - i-jmr_v15i1e96779_app3.docx ]

## References

1. National Academy of Medicine. In: Whicher D, Ahmed M, Israni ST, Matheny M, editors. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril: National Academies Press (US); 2023. [doi: 10.17226/27111]
2. Triantafyllidis AK, Tsanas A. Applications of machine learning in real-life digital health interventions: review of the literature. J Med Internet Res 2019 Apr 5;21(4):e12286. [doi: 10.2196/12286] [Medline: 30950797]
3. Russell SJ, Norvig P. Artificial Intelligence: A Modern Approach: Prentice Hall; 1995.
4. Witten IH, Frank E, Hall MA, Pal CJ. Data Mining: Practical Machine Learning Tools and Techniques: Morgan Kaufmann; 2017. [doi: 10.1016/C2015-0-02071-8]
5. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. Lancet 2020 May 16;395(10236):1579-1586. [doi: 10.1016/S0140-6736(20)30226-9] [Medline: 32416782]
6. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. J Med Internet Res 2019 Jul 10;21(7):e13659. [doi: 10.2196/13659] [Medline: 31293245]
7. Vishnoi S, Yadav SL, Saini AK, Vats P. Transforming healthcare delivery: assessing the role of e-governance strategies in IOMT and AIML-based medical system administration during the pandemic in the Indian subcontinent. Biomed Pharmacol J 2025;18(December Spl Edition):331-346. [doi: 10.13005/bpj/3091]
8. Kaladharan S, Manayath D, Gopalakrishnan R. Regulatory challenges in AI/ML-enabled medical devices: a scoping review and conceptual framework. J Med Device 2024 Dec;18(4):040801. [doi: 10.1115/1.4066280]
9. Basulo-Ribeiro J, Teixeira L. The future of healthcare with industry 5.0: preliminary interview-based qualitative analysis. Future Internet 2024;16(3):68. [doi: 10.3390/fi16030068]
10. Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. J R Soc Med 2019 Jan;112(1):22-28. [doi: 10.1177/0141076818815510] [Medline: 30507284]
11. Saha E, Rathore P. The impact of healthcare 4.0 technologies on healthcare supply chain performance: extending the organizational information processing theory. Technol Forecast Soc Change 2024 Apr;201:123256. [doi: 10.1016/j.techfore.2024.123256]
12. Collins GS, Moons KG. Reporting of artificial intelligence prediction models. Lancet 2019 Apr 20;393(10181):1577-1579. [doi: 10.1016/S0140-6736(19)30037-6] [Medline: 31007185]
13. Hazarika I. Artificial intelligence: opportunities and implications for the health workforce. Int Health 2020 Jul 1;12(4):241-245. [doi: 10.1093/inthealth/ihaa007] [Medline: 32300794]
14. Lungu DA, Grillo Ruggieri T, Nuti S. Decision making tools for managing waiting times and treatment rates in elective surgery. BMC Health Serv Res 2019 Jun 11;19(1):369. [doi: 10.1186/s12913-019-4199-6] [Medline: 31185989]
15. Agboola S, Hale TM, Masters C, Kvedar J, Jethwani K. "Real-world" practical evaluation strategies: a review of telehealth evaluation. JMIR Res Protoc 2014 Dec 17;3(4):e75. [doi: 10.2196/resprot.3459] [Medline: 25524892]
16. Peters MD, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Scoping reviews. In: Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, editors. JBI Manual for Evidence Synthesis: JBI; 2024. [doi: 10.46658/JBIMES-24-09]



XSL•FO
RenderX

17. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
18. Rocha DM, Barbosa GV, De Souza Bermejo PH. Application of machine learning algorithms in health systems planning and public policy: a scoping review. Open Science Framework. Preprint posted online on Jul 10, 2026. [doi: 10.17605/OSF.IO/VWBHX]
19. Rocha DMA. Search strategy about applications of machine learning algorithms in the context of real-life interventions and resulting clinical implications to guide planning and public policy formulation actions for the health system (MEDLINE). searchRxiv. Preprint posted online on Sep 18, 2025. [doi: 10.1079/searchRxiv.2025.01103]
20. Rocha DM. Search strategy about applications of machine learning algorithms in the context of real-life interventions and resulting clinical implications to guide planning and public policy formulation actions for the health system (EMBASE). searchRxiv. Preprint posted online on Sep 18, 2025. [doi: 10.1079/searchRxiv.2025.01104]
21. Rocha DM. Search strategy about applications of machine learning algorithms in the context of real-life interventions and resulting clinical implications to guide planning and public policy formulation actions for the health system (Scopus). searchRxiv. Preprint posted online on Sep 18, 2025. [doi: 10.1079/searchRxiv.2025.01105]
22. Rocha DM. Search strategy about applications of machine learning algorithms in the context of real-life interventions and resulting clinical implications to guide planning and public policy formulation actions for the health system (LILACS). searchRxiv. Preprint posted online on Sep 18, 2025. [doi: 10.1079/searchRxiv.2025.01106]
23. Campagner A, Berjano P, Lamartina C, Langella F, Lombardi G, Cabitza F. Assessment and prediction of spine surgery invasiveness with machine learning techniques. Comput Biol Med 2020 Jun;121:103796. [doi: 10.1016/j.compbiomed.2020.103796] [Medline: 32568677]
24. Ye C, Wang O, Liu M, et al. A real-time early warning system for monitoring inpatient mortality risk: prospective study using electronic medical record data. J Med Internet Res 2019 Jul 5;21(7):e13719. [doi: 10.2196/13719] [Medline: 31278734]
25. Fontana MA, Lyman S, Sarker GK, Padgett DE, MacLean CH. Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty? Clin Orthop Relat Res 2019 Jun;477(6):1267-1279. [doi: 10.1097/CORR.0000000000000687] [Medline: 31094833]
26. Karhade AV, Ogink P, Thio Q, et al. Development of machine learning algorithms for prediction of discharge disposition after elective inpatient surgery for lumbar degenerative disc disorders. Neurosurg Focus 2018 Nov 1;45(5):E6. [doi: 10.3171/2018.8.FOCUS18340] [Medline: 30453463]
27. Hayn D, Kreiner K, Ebner H, et al. Development of multivariable models to predict and benchmark transfusion in elective surgery supporting patient blood management. Appl Clin Inform 2017 Jun 14;8(2):617-631. [doi: 10.4338/ACI-2016-11-RA-0195] [Medline: 28850152]
28. Krämer J, Schreyögg J, Busse R. Classification of hospital admissions into emergency and elective care: a machine learning approach. Health Care Manag Sci 2019 Mar;22(1):85-105. [doi: 10.1007/s10729-017-9423-5] [Medline: 29177993]
29. EndNote. Clarivate. URL: https://clarivate.com/accessibility/endnote/ [accessed 2026-08-23]
30. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210. [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
31. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015 Jan 6;162(1):55-63. [doi: 10.7326/M14-0697] [Medline: 25560714]
32. EPHPP quality assessment tool for quantitative studies 2026. Elite Providers Hub for Progressive Play. URL: https://www.ephpp.ca/quality-assessment-tool-for-quantitative-studies/ [accessed 2026-08-23]
33. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. J Eval Clin Pract 2012 Feb;18(1):12-18. [doi: 10.1111/j.1365-2753.2010.01516.x] [Medline: 20698919]
34. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
35. Stevens RJ, Poppe KK. Validation of clinical prediction models: what does the “calibration slope” really measure? J Clin Epidemiol 2020 Feb;118:93-99. [doi: 10.1016/j.jclinepi.2019.09.016] [Medline: 31605731]
36. Toma TS, Pereira TV, Vanni T, Barreto JO. Avaliac¸a˜o de tecnologias de sau´de & poli´ticas informadas por evide^ncias [Book in Portuguese]: Instituto de Saúde; 2017.
37. Ramezani M, Mobinizadeh M, Bakhtiari A, et al. Agenda setting for health equity assessment through the lenses of social determinants of health using machine learning approach: a framework and preliminary pilot study. BioData Min 2025 Feb 10;18(1):14. [doi: 10.1186/s13040-025-00428-x] [Medline: 39930525]
38. E-SUS primary health care strategy (e-SUS APS). Ministry of Health, Brasil. 2021. URL: https://www.gov.br/saude/pt-br/assuntos/esua-aps [accessed 2026-10-29]
39. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive




XSL•FO
RenderX

95/46/EC (General Data Protection Regulation) (Text with EEA relevance). European Union. 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng [accessed 2026-10-29]
40. General Personal Data Protection Law (LGPD) - law no. 13.709/2018. Federal Government of Brazil. 2018. URL: https://www2.senado.leg.br/bdsf/handle/id/545155 [accessed 2026-10-29]
41. Collins GS, Moons KG, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024 Apr 16;385:e078378. [doi: 10.1136/bmj-2023-078378] [Medline: 38626948]
42. Moons KG, Damen JA, Kaul T, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ 2025 Mar 24;388:e082505. [doi: 10.1136/bmj-2024-082505] [Medline: 40127903]
43. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 2016 Mar 15;3:160018. [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
44. Porter ME. A strategy for health care reform--toward a value-based system. N Engl J Med 2009 Jul 9;361(2):109-112. [doi: 10.1056/NEJMp0904131] [Medline: 19494209]
45. Porter ME. What is value in health care? N Engl J Med 2010 Dec 23;363(26):2477-2481. [doi: 10.1056/NEJMp1011024] [Medline: 21142528]
46. Qinggui H, Xiuqiong T, Hui C, Jinsong W. The ethical risks and solutions of the algorithm black box in medical artificial intelligence. Acta Bioeth 2025;31(1):24-34. [doi: 10.4067/S1726-569X2025000100024]
47. Huo Z, Booth J, Monks T, et al. Dynamic mortality prediction in critically ill children during interhospital transports to PICUs using explainable AI. NPJ Digit Med 2025 Feb 17;8(1):108. [doi: 10.1038/s41746-025-01465-w] [Medline: 39962177]
48. Gu F, Meyer AJ, Ježek F, et al. Identification of digital twins to guide interpretable AI for diagnosis and prognosis in heart failure. NPJ Digit Med 2025 Feb 18;8(1):110. [doi: 10.1038/s41746-025-01501-9] [Medline: 39966509]
49. Cammarata R, Ruffini F, Catamerò A, et al. Machine learning for predicting the low risk of postoperative pancreatic fistula after pancreaticoduodenectomy: toward a dynamic and personalized postoperative management strategy. Cancers (Basel) 2025 May 31;17(11):1846. [doi: 10.3390/cancers17111846] [Medline: 40507327]
50. Singh NP, Mujawar MA, Golani A. Role of artificial intelligence in enhancing mechanical ventilation - a peek into the future. Indian J Anaesth 2025 Jul;69(7):722-728. [doi: 10.4103/ija.ija_995_24] [Medline: 40657139]

## Abbreviations

**ANN:** artificial neural network
**AUC:** area under the curve
**AUROC:** area under the receiver operating characteristic curve
**CDSS:** clinical decision support system
**EPHPP:** Effective Public Health Practice Project
**FAIR:** findable, accessible, interoperable, and reusable
**GDP:** gross domestic product
**GDPR:** General Data Protection Regulation
**H5.0:** Health 5.0
**JBI:** Joanna Briggs Institute
**LASSO:** least absolute shrinkage and selection operator
**LGPD:** Lei Geral de Proteção de Dados Pessoais
**LIME:** local interpretable model-agnostic explanations
**MCID:** minimal clinically important difference
**ML:** machine learning
**NASSS:** nonadoption, abandonment, and scale-up, spread, and sustainability
**NLP:** natural language processing
**PHC:** primary health care
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
**PROBAST:** Prediction Model Risk of Bias Assessment Tool
**PROM:** patient-reported outcome measure
**RF:** random forest
**RQ:** research question
**RWD:** real-world data
**SDOH:** social determinants of health
**SHAP:** Shapley additive explanations
**SUS:** Sistema Único de Saúde



XSL•FO
RenderX

**SVM:** support vector machine
**TRIPOD:** Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
**XAI:** explainable AI

*Edited by M Balcarras; submitted 20.Apr.2026; peer-reviewed by CN Herrera, FA Causio, JC Boucher, J Odeniya; revised version received 01.Sep.2026; accepted 14.Sep.2026; published 16.Nov.2026.*

*Please cite as:*
*Rocha DMA, Barbosa GV, De Souza Bermejo PH*
*Application of Machine Learning Algorithms to Health Systems Planning and Public Policy Using Real-World Data: Scoping Review*
*Interact J Med Res 2026;15:e96779*
*URL: https://www.i-jmr.org/2026/1/e96779*
*doi:10.2196/96779*

© Deusa Maria Arruda Rocha, Gentil Veloso Barbosa, Paulo Henrique De Souza Bermejo. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 16.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Impact of Elevated Blood Pressure Targets on Splanchnic Ischemia and Mortality in Septic Resuscitation: Systematic Review

Paulina Cisneros[1], MD; John Fredy Ortiz Caviedes[2], MD; Lina Maria Piñeros Ladino[3], MD; Natalia Quintero[4], MD; Valeria García Oñate Cervera[5]; Juan David Urrea Ospina[6]; Erika Patricia Cerna Alcantar[7], MD

[1]Cardiology Department, Guthrie Forssmann Cardiovascular Institute, Hospital Enrique Garcés, Chilibulo S/N y Enrique Garcés, Quito, Ecuador

[2]Internal Medicine Department, Hospital San Rafael, Bogota, Colombia

[3]Medicine Department, Preventivamente Salud IPS, Medellín, Colombia

[4]Clinical Toxicology Department, Fundación Universitaria de Ciencias de la Salud FUCS, Bogotá, Colombia

[5]Medical Department, Escuela Medico Militar, Ciudad de México, Mexico

[6]Medical Department, Universidad Libre de Colombia, Cali, Colombia

[7]Internal Medicine Department, Mexican Social Security Institute, San Luis Potosí, Mexico

**Corresponding Author:**
Paulina Cisneros, MD
Cardiology Department, Guthrie Forssmann Cardiovascular Institute, Hospital Enrique Garcés, Chilibulo S/N y Enrique Garcés, Quito, Ecuador

## Abstract

**Background:** Sepsis causes deaths worldwide each day, with hemodynamic resuscitation being central to management. Guidelines recommend a 65-mm Hg mean arterial pressure (MAP) target, yet the evidence for elevated targets remains fragmented across mortality, renal, and splanchnic outcomes.

**Objective:** This review aims to comprehensively appraise the impact of elevated MAP targets on splanchnic ischemia, mortality, and organ-specific outcomes in septic resuscitation.

**Methods:** Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines, we systematically searched PubMed, the Cochrane Library, and Google Scholar for studies comparing MAP targets in septic shock. Risk of bias was assessed using the Cochrane Risk of Bias 2 (RoB 2; randomized controlled trials [RCTs]) and the Newcastle-Ottawa Scale (nonrandomized studies). The findings were synthesized narratively according to the Synthesis Without Meta-Analysis guidelines, with random-effects meta-analysis for mortality outcomes. Evidence certainty was graded using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) framework.

**Results:** From 3254 records, 11 studies (2895 patients; 4 RCTs, 7 nonrandomized) met the inclusion criteria. Pooled 90-day mortality across 3 trials showed a nonsignificant trend toward higher mortality with elevated MAP targets (320/720, 44.4% vs 280/722, 38.8%; OR 1.31, 95% CI 0.97 - 1.77). Pooled 28- or 30-day mortality showed a similar nonsignificant trend (OR 1.25, 95% CI 0.98 - 1.61), which was robust to sensitivity analysis. Narratively, mortality findings across nonpooled observational studies were inconsistent; some linked low MAP to higher death risk, and others found no significant association. Renal outcomes were inconsistent, with some studies showing reduced renal replacement therapy or improved acute kidney injury recovery at higher targets, particularly in hypertensive or cirrhotic subgroups. Elevated MAP targets consistently increased vasopressor exposure and were associated with more arrhythmias and fewer organ support–free days. No study reported splanchnic ischemia as a discrete outcome; mesenteric ischemia rates were low and similar between arms. Overall certainty of evidence was very low.

**Conclusions:** Higher MAP targets show no clear mortality or splanchnic perfusion benefit and may increase overall harm, though select subgroups, chronic hypertension, and cirrhosis, may benefit, supporting individualized, risk-stratified MAP targeting over uniformly elevated-target strategies.

***(Interact J Med Res 2026;15:e98772)*** doi:10.2196/98772

**KEYWORDS**

septic shock; mean arterial pressure; vasopressor therapy; hemodynamic resuscitation; splanchnic perfusion; acute kidney injury



XSL•FO
RenderX

## Introduction

Sepsis affected an estimated 48.9 million people and caused 11 million deaths worldwide in 2017 (nearly 1 in 5 deaths globally), equating to more than 30,000 deaths per day [1]. Globally, sepsis incidence rose by 230% among individuals aged 15 years and older between 1990 and 2021, with a further 30% surge during 2020 - 2021 alone amid the COVID-19 pandemic, underscoring an urgent, sustained need for advances in sepsis therapy [2]. Due to socioeconomic difficulties and a lack of thorough epidemiological surveys, sepsis mortality continues to exist in low-income countries [3].

Maintaining a mean arterial pressure (MAP) of 65 mm Hg is advised by current sepsis management recommendations, which state, as a strong recommendation based on moderate-quality evidence, that there is no appreciable additional benefit to targeting a higher MAP in patients with septic shock [4]. An MAP of more than 70 mm Hg has been shown in several trials to improve microcirculation during septic shock and to avoid acute renal damage [5,6]. In patients with sepsis, hemodynamic control is an essential part of the overall therapeutic strategy [7], as it plays a critical role in early vital sign stabilization, anticipating disease progression, and avoiding further tissue and organ failure [8]. However, maintaining a high MAP may need higher doses of vasoactive medications, which might result in further issues and bodily harm [9].

Despite guideline endorsement of a 65-mm Hg MAP target in septic shock, the evidence base for elevated MAP targets remains fragmented and outdated. Leone et al [8] synthesized only studies through 2014, predating 3 pivotal trials, with the latter halted early for harm, leaving current practice uninformed by the most recent and arguably most consequential data. Dari et al [10], while more recent, pooled only 4 randomized controlled trials (RCTs) and excluded observational and cohort evidence entirely, yielding a narrower and potentially incomplete effect estimate. A study by Kattan et al [11] offered valuable mechanistic and monitoring insight but remained a narrative review, lacking quantitative synthesis or a dedicated, pooled assessment of splanchnic and mesenteric outcomes. Critically, no prior review has directly synthesized splanchnic or mesenteric ischemia as a discrete, cross-study outcome nor systematically reconciled short- and long-term mortality signals alongside renal, hemodynamic, and organ-support end points across both randomized and observational designs. Our review addresses these gaps by incorporating the most contemporary RCTs and observational cohorts, applying sensitivity analyses to resolve heterogeneity, and explicitly evaluating splanchnic perfusion alongside mortality. The aim of this review is to comprehensively and quantitatively appraise the impact of elevated MAP targets on splanchnic ischemia, mortality, and organ-specific outcomes in septic resuscitation.

## Methods

### Study Design

This systematic review was conducted and reported in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines and checklist (Checklist 1) [12]. The review protocol was developed a priori, with the research question, eligibility criteria, search strategy, and outcome hierarchy defined before data extraction commenced. The review was not registered prospectively on PROSPERO, which is acknowledged as a limitation.

### Eligibility Criteria

#### *Inclusion Criteria*

The inclusion criteria were as follows:

- Adult patients (≥18 y) diagnosed with sepsis, severe sepsis, or septic shock according to Sepsis-2 or Sepsis-3 consensus definitions.
- Patients admitted to the intensive care unit (ICU) requiring vasopressor-supported hemodynamic resuscitation.
- Studies involving septic shock in patients with underlying comorbidities (eg, cirrhosis, chronic kidney disease, or chronic hypertension) were eligible provided that MAP-related clinical outcomes were reported separately.
- RCTs, quasi-randomized trials, prospective interventional studies, post hoc analyses of clinical trials, and prospective or retrospective observational cohort studies.
- Studies evaluating different MAP targets or achieved MAP levels during septic shock resuscitation, including higher MAP targets (generally ≥75 - 85 mm Hg), standard MAP targets (generally 60 - 70 mm Hg), or analyses of MAP as a continuous or categorical exposure variable.
- Studies reporting at least one of the following outcomes:
  - Primary outcomes: all-cause mortality or splanchnic perfusion- or ischemia-related outcomes.
  - Secondary outcomes: acute kidney injury (AKI), renal replacement therapy (RRT), organ dysfunction (eg, Sequential Organ Failure Assessment [SOFA] score), lactate, hemodynamic parameters, vasopressor requirements, mechanical ventilation, ICU or hospital length of stay, or adverse events.

#### *Exclusion Criteria*

The exclusion criteria were as follows:

- Studies involving nonseptic populations, or mixed populations in which data for patients with septic shock could not be extracted separately.
- Pediatric populations (<18 y).
- Studies evaluating vasoactive therapy without reporting MAP targets or achieved MAP levels.
- Case reports, case series with fewer than 10 patients, editorials, letters, narrative reviews, conference abstracts without full text, protocols, animal studies, and in vitro studies.
- Studies not reporting any predefined primary or secondary clinical outcome relevant to this review.
- Duplicate publications or secondary analyses of the same dataset, unless they reported additional relevant outcomes not available in the primary publication.

### Information Sources and Search Strategy

A comprehensive systematic literature search was conducted across 3 electronic bibliographic databases: MEDLINE (via



XSL•FO
RenderX

PubMed), Google Scholar, and the Cochrane Library. No restrictions were placed on publication date or language during the initial search. However, due to resource limitations, only studies with full-text articles available in English were included in the final review, which may have introduced language bias.

The search strategy was developed using a combination of MeSH and free-text keywords based on the Population, Intervention, Comparator, and Outcome (PICO) framework. For PubMed, the following search strategy was applied: ("Sepsis"[Mesh] OR "Shock, Septic"[Mesh] OR sepsis OR "septic shock" OR "severe sepsis") AND ("Blood Pressure"[Mesh] OR "Mean Arterial Pressure" OR MAP OR "mean arterial pressure" OR "blood pressure target" OR "higher MAP" OR "high MAP" OR "elevated blood pressure" OR "hemodynamic resuscitation" *OR vasopressor* OR norepinephrine OR vasopressin) AND ("Splanchnic Circulation"[Mesh] OR "Mesenteric Ischemia"[Mesh] OR splanchnic OR "splanchnic perfusion" OR "splanchnic ischemia" OR mesenteric OR gut OR intestinal OR gastrointestinal OR gastric OR hepatic OR liver OR renal) AND ("Mortality"[Mesh] OR mortality OR death OR survival OR "Organ Dysfunction Scores"[Mesh] OR "organ dysfunction" OR "multiple organ failure" OR "acute kidney injury" OR AKI OR lactate).

For the Cochrane Central Register of Controlled Trials (CENTRAL), an adapted search strategy using free-text terms was applied: (sepsis OR "septic shock" OR "severe sepsis") AND ("mean arterial pressure" OR MAP OR "blood pressure target" OR "higher MAP" OR "high MAP" OR "elevated blood pressure" OR "hemodynamic resuscitation" OR vasopressor OR norepinephrine OR vasopressin) AND (splanchnic OR "splanchnic perfusion" OR "splanchnic ischemia" OR mesenteric OR intestinal OR gastrointestinal OR gastric OR hepatic OR liver OR renal) AND (mortality OR death OR survival OR "organ dysfunction" OR "multiple organ failure" OR "acute kidney injury" OR AKI OR lactate).

Database-specific syntax and Boolean operators were adapted as appropriate for each platform. For Google Scholar, a simplified combination of key search terms based on the PICO framework was used, and the first 400 results, sorted by relevance, were screened because of the platform's limitation in exporting large numbers of citations. In addition, the reference lists of all included studies and relevant review articles were manually searched to identify any additional eligible studies that were not retrieved through the electronic database searches.

## Study Selection

### *Overview*

Search results from all databases were collated and deduplicated using reference management software. Two independent reviewers screened titles and abstracts against the prespecified eligibility criteria. Studies judged potentially eligible at the abstract screening stage were retrieved as full-text articles and subjected to detailed eligibility assessment. Disagreements between reviewers at any stage of screening were resolved through discussion and, where necessary, by referral to a third reviewer. The reasons for exclusion at the full-text stage were documented and are summarized in the PRISMA flow diagram. Pettilä [13] et al (TARTARE-2S, 2026) was excluded as it randomized a composite tissue-perfusion-guided protocol (capillary refill, skin mottling, temperature, lactate) rather than a fixed or isolated MAP target, with MAP serving only as a subordinate safety limit.

### *Data Extraction*

Data were extracted independently by 2 reviewers using a standardized, prepiloted extraction form, with disagreements resolved through discussion and, when necessary, consultation with a third reviewer. Extracted variables included study characteristics (design, setting, and sample size), participant demographics and baseline clinical characteristics, relevant comorbidities (eg, chronic hypertension, chronic kidney disease, and cirrhosis), MAP intervention targets or observational exposure measures, vasopressor type and dose, and reported outcomes. Primary outcomes comprised all-cause mortality at the reported follow-up time points and splanchnic perfusion parameters. Secondary outcomes included AKI, RRT, serum creatinine, urine output, hemodynamic variables, vasopressor requirements, organ support–free days, SOFA scores, length of stay, and adverse events. Where applicable, subgroup analyses reported within individual studies were also extracted. Data presented only in graphical format were estimated from figures, and unpublished data were not requested from study investigators.

### *Risk of Bias Assessment*

Methodological quality was assessed independently by 2 reviewers using validated tools according to study design, with disagreements resolved by consensus. RCTs were evaluated using the Cochrane Risk of Bias 2 (RoB 2) tool, which assesses bias arising from the randomization process, deviations from intended interventions, missing outcome data, outcome measurement, and selection of the reported results [14]. Nonrandomized studies were assessed using the Newcastle-Ottawa Scale (NOS), which evaluates study quality across the domains of selection, comparability, and outcome assessment, with a maximum score of 9 points. Overall study quality was categorized according to the respective assessment tools [15].

The remaining nonrandomized study involving 2 actively assigned MAP-targeted groups was assessed using the Risk of Bias in Non-randomized Studies of Interventions (ROBINS-I) tool, evaluating bias due to confounding, selection of participants, classification of interventions, deviations from intended interventions, missing data, outcome measurement, and selection of the reported result [16]. The single-arm, within-subject titration study lacking a comparator group was assessed using the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Quasi-Experimental Studies, a 9-item tool suited to pre-post designs without a control arm [17]. Overall quality was categorized per the respective tool applied, and studies were synthesized with attention to how these quality differences might affect confidence in the findings.



XSL•FO
RenderX

### *Data Synthesis*

Given the substantial clinical and methodological heterogeneity among the included studies, the findings were synthesized narratively in accordance with the Synthesis Without Meta-Analysis reporting guideline [18]. Studies were grouped according to study design and outcome domains, with evidence from RCTs presented before nonrandomized studies. Quantitative meta-analysis was performed only for all-cause mortality because sufficient comparable interventional data were available. Separate random-effects meta-analyses were conducted for 90-day mortality and 28-day mortality, with an additional sensitivity analysis incorporating studies reporting analogous 30-day mortality. Pooled effect estimates were expressed as odds ratios (ORs) with 95% CIs, and statistical heterogeneity was assessed using Cochran *Q* test and the $I^2$ statistic. Meta-analysis was not undertaken for renal, hemodynamic, organ-support, or adverse event outcomes because of a lack of reported statistics, differences in study populations, MAP targets, intervention protocols, follow-up durations, and outcome definitions. Subgroup findings reported within individual studies were synthesized narratively but were not quantitatively pooled because they were primarily exploratory or post hoc analyses.

### *Certainty of Evidence*

The overall certainty of the evidence for the primary outcomes was assessed using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) framework [19]. Evidence from RCTs was initially considered of high certainty and downgraded, where appropriate, for risk of bias, inconsistency, indirectness, imprecision, and publication bias. Evidence from nonrandomized studies was initially considered of low certainty and was upgraded only when justified by large effect estimates, evidence of a dose-response relationship, or when residual confounding was unlikely to explain the observed findings. Publication bias was assessed qualitatively given the small number of studies (n=3 - 4) available for each pooled outcome, which precluded the reliable use of funnel plot asymmetry testing or Egger regression; this qualitative assessment considered the comprehensiveness of the search strategy, inclusion of published literature identified through multiple databases and manual reference screening, and comparison of effect directions and magnitudes across studies of differing design and setting, with no evidence suggestive of publication bias (eg, selective reporting or a disproportionate predominance of positive findings) identified. The overall certainty of evidence for each primary outcome was classified as high, moderate, low, or very low and is reported alongside the corresponding results.

## Results

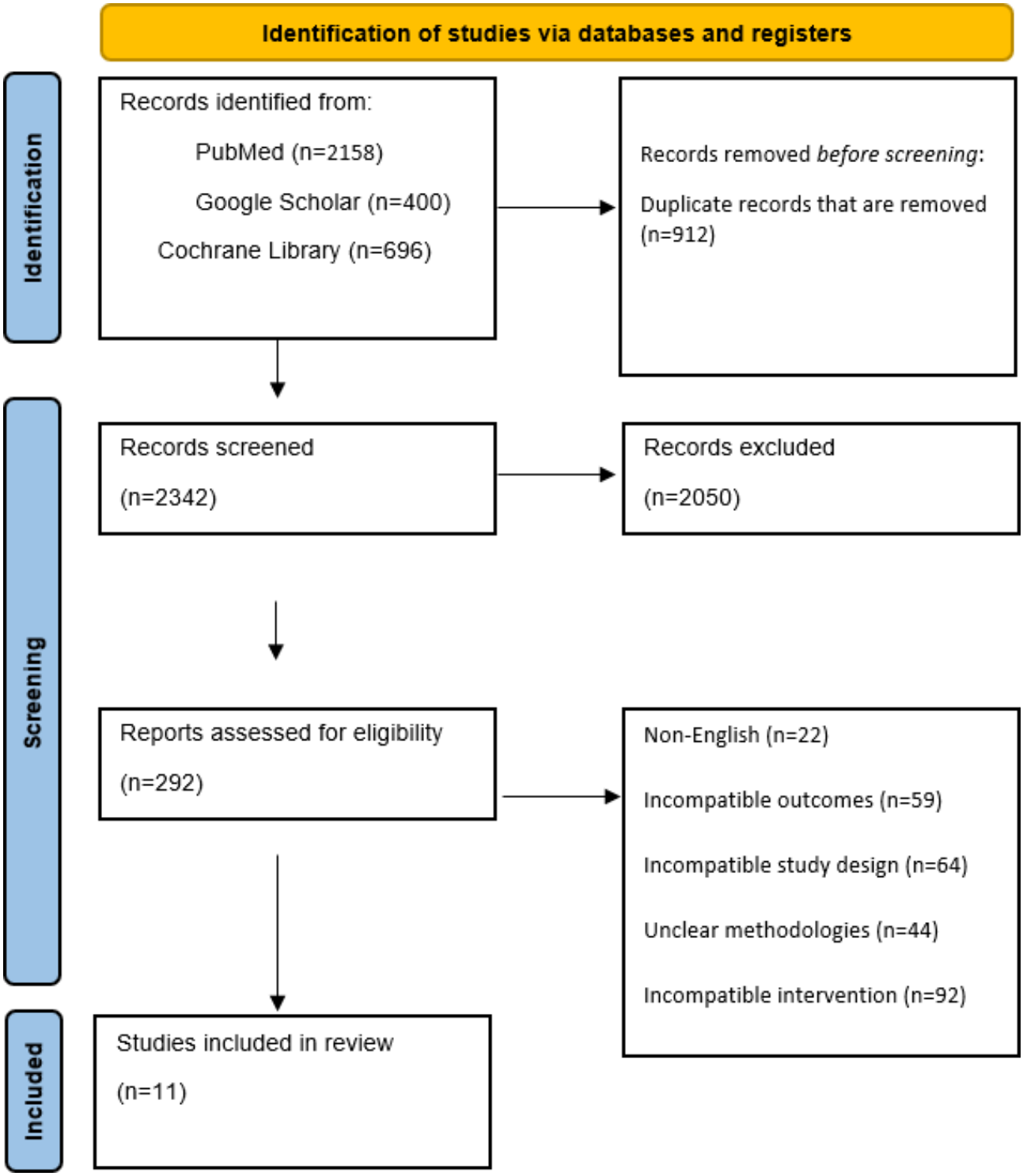
### Study Selection

A systematic search was conducted across PubMed or MEDLINE, the Cochrane Library, and Google Scholar from inception through April 2026, using predefined search terms related to MAP targets, septic shock, splanchnic ischemia, and resuscitation outcomes. A total of 3254 records were identified through database searching and manual reference screening, of which 912 duplicates were removed. Following title and abstract screening, 292 full-text articles were assessed for eligibility against the prespecified inclusion criteria. Studies were excluded due to nonseptic populations, absence of a discrete MAP target comparison, insufficient outcome reporting, or nonhuman design. Ultimately, 11 studies met all inclusion criteria and were incorporated into the qualitative synthesis. The full selection process is summarized in the PRISMA flow diagram (Figure 1).



XSL•FO
RenderX

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.




## Study Characteristics

### *Overview*

Eleven studies (2895 patients) examined MAP targets in septic shock, including 4 RCTs and 7 nonrandomized studies (including prospective and retrospective observational cohorts, a post hoc analysis, and a physiological crossover study), with sample sizes ranging from 10 to 776. Populations were generally older adults with high illness severity (Acute Physiology and Chronic Health Evaluation II [APACHE II] 15 - 29, Simplified Acute Physiology Score II [SAPS II] 38 - 58, SOFA 7 - 12.5), although age ranged from younger patients with cirrhosis (~45 y; Maiwall et al [20]) to an older adults-only cohort (≥65 y). Four studies compared fixed high- versus low-MAP targets (typically 80 - 85 mm Hg vs 65 - 70 mm Hg, except TARGET-C, which compared 80 - 85 mm Hg vs 60 - 65 mm Hg), whereas the remaining studies evaluated MAP as a continuous exposure or through individualized targets rather than fixed intervention arms. Comorbidities varied across cohorts, with chronic hypertension reported in 9% to 53% of patients, alongside chronic kidney disease, diabetes, coronary artery disease, and, in 1 RCT, cirrhosis with AKI (Table 1).



XSL•FO
RenderX

**Table .** Characteristics and results of the studies included.

| Study | Design | Population (N) | Age (y) | Male (%) | Baseline severity | MAP[a] comparison (high vs low or target) | Key comorbidities or notes |
|---|---|---|---|---|---|---|---|
| Bourgoin et al [21] (2005) | Single-center, open-label RCT[b] (parallel-group, run-in stabilization) | Severe septic shock; N=28 (14/14) | G1: 65 (median); G2: 63 (median) | G1: 10/14 (71%); G2: 9/14 (64%) | APACHE II[c]: 27 vs 28 | 85 mm Hg vs 65 mm Hg (4-hour exposure) | Dopamine excluded; single vasopressor (norepinephrine) protocol |
| Badin et al [22] (2011) | Prospective observational cohort, 2 ICUs[d], nonrandomized | Sustained hypotension, mixed shock etiologies; N=217 (septic shock subgroup n=127, 59%) | Mean 64 (SD 15) | 127/217 (59%) | SAPS II: mean 53.2 (SD 18) | MAP as continuous exposure (no fixed arms) | Chronic HTN[e] 41%; type 2 DM[f] 15%; institutional MAP target >65 mm Hg |
| Poukkanen et al [23] (2013; FINNAKI[g]) | Prospective multicenter observational substudy, 17 ICUs, nonrandomized | Severe sepsis with continuous hemodynamic data; N=423 | ~63 - 64 (median) | ~63% overall | SAPS II[h]: 38 - 43; SOFA[i] D1: 7 - 9 | MAP as time-adjusted continuous exposure (no fixed arms) | HTN ~48% - 53%; CKD[j] 11.1% vs 2.6%; DM 27.5% vs 18.5% |
| Endo et al [24] (2025; OPT-PRESS[k]) | Multicenter, open-label RCT, Japan; stopped early for harm | Septic shock, age ≥65 y; N=516 (257/259) | High: 79 (IQR 74 - 84); control: 77 (IQR 73 - 85) | High: 143/257 (56%); control: 137/259 (53%) | SOFA: 9 - 10; APACHE II: 25 - 27 | 80 - 85 mm Hg vs 65 - 70 mm Hg | Chronic HTN ~53% (stratified variable); 97% - 99% Japanese cohort |
| Asfar et al [25] (2014; SEPSISPAM)[l] | Multicenter (29 centers, France), open-label RCT, stratified by chronic HTN | Septic shock refractory to fluids; N=776 (388/388) | High: mean 65 (SD 13); low: mean 65 (SD 15) | Not separately reported | SAPS II:~56 - 57; SOFA:~10.7 - 10.8 | 80 - 85 mm Hg vs 65 - 70 mm Hg | Chronic HTN ~44%; CKD 6% - 8%; cirrhosis ~7% |
| Maiwall et al [20] (2023; TARGET-C[m]) | Single-center RCT (India), block randomization | Cirrhosis + septic shock + AKI[n]; N=150 (75/75) | Low: mean 46.7 (SD 9.1); high: mean 45.0 (SD 12.9) | Low: 65/75 (87%); high: 68/75 (91%) | SOFA: ~12; MELD:~31.5 - 32.5 | 80 - 85 mm Hg vs 60 - 65 mm Hg | All patients cirrhotic; HTN 9% both arms |
| LeDoux et al [26] (2000) | Single-center, within-patient crossover or titration study | Septic shock; N=10 | Mean 68 (SD 12) | 7/10 (70%) | APACHE II: 29 | Sequential 65→75→85 mm Hg (no separate control arm) | Only 3/10 had prior HTN; dopamine used as co-vasoactive in 5/10 (confounder) |
| Dünser et al [27] (2009; Crit Care [546C88 post hoc]) | Post hoc regression analysis of RCT placebo arm | Septic shock, control/placebo group; N=290 included (68 excluded, MAP <70) | Mean 60 (SD 15) | 178/290 (61%) | SAPS II: 48 - 50 | MAP quartiles (observed exposure, not randomized) | Chronic HTN 31.4%; MV[o] required in 96.2% |
| Kuragayala et al [28] (2019) | Prospective observational, single-center (India), 2 MAP groups | Septic shock refractory to fluids; N=100 (50/50) | Low: mean 58.56 (SD 18.25); high: mean 57.29 (SD 17) | Low: 28/50 (56%); high: 27/50 (54%) | SAPS II:~57 - 58; SOFA:~10.4 - 10.5 | Target 80 - 85 mm Hg vs 65 - 70 mm Hg (achieved: mean 78.45 (SD 17.23) vs mean 57.32 (SD 13.69) | Chronic HTN ~19% both arms; no inferential statistics reported |
| Varpula et al [29] (2005) | Retrospective single-center cohort | Septic shock requiring vasopressors; N=111 (75 survivors or 36 nonsurvivors) | Survivors: mean 48.8 (SD 15.5); nonsurvivors: mean 55.3 (SD 14.6) | Survivors: 44/75 (59%); nonsurvivors: 18/36 (50%) | APACHE II: 15 vs 21; SOFA D1: 8 vs 12 | MAP as continuous exposure (institutional target >65 mm Hg, no arms) | Positive blood culture ~55%; pneumonia most common source |



XSL•FO
RenderX

| Study | Design | Population (N) | Age (y) | Male (%) | Baseline severity | MAP[a] comparison (high vs low or target) | Key comorbidities or notes |
|---|---|---|---|---|---|---|---|
| Dünser et al [30] (2009; Bern cohort) | Retrospective single-center cohort | Sepsis or septic shock; N=274 (sepsis 126; septic shock 148) | mean 61 (SD 16) | 170/274 (62%) | SOFA: mean 10.6 (SD 5.1); SAPS II: mean 51.5 (SD 19.9); APACHE II: mean 27.3 (SD 7.8) | Individualized MAP targets 50 - 75 mm Hg (continuous exposure, no fixed arms) | Chronic HTN 38%; coronary heart disease 22.6% |

[a]MAP: mean arterial pressure.

[b]RCT: randomized controlled trial.

[c]APACHE II: Acute Physiology and Chronic Health Evaluation II.

[d]ICU: intensive care unit.

[e]HTN: hypertension.

[f]DM: diabetes mellitus.

[g]FINNAKI: Finnish Acute Kidney Injury Study *(study acronym).*

[h]SAPS II; Simplified Acute Physiology Score II.

[i]SOFA: Sequential Organ Failure Assessment.

[j]CKD: chronic kidney disease.

[k]OPTPRESS: Optimal Blood Pressure in Septic Shock *(study acronym).*

[l]SEPSISPAM: Sepsis and Mean Arterial Pressure *(study acronym).*

[m]TARGET-C: Targeted Mean Arterial Pressure in Cirrhosis with Septic Shock *(study acronym).*

[n]AKI: acute kidney injury.

[o]MV: mechanical ventilation.

### *Quality Assessment*

Among the 4 RCTs, 2 studies were judged to have a low overall risk of bias [24,25], whereas Bourgoin et al [21] and Maiwall et al [20] were rated as having some concerns. The main limitations were related to deviations from intended interventions, missing outcome data, and outcome measurement in the open-label trials (Figure 2).

The methodological quality of the 7 nonrandomized studies ranged from very low to good. NOS scores varied between 2 and 8, with Poukkanen et al [23] demonstrating the highest quality (8/9). Badin et al [22], Varpula et al [29], Dünser et al [27], and Dünser et al [30] achieved moderate-to-good quality (7/9; Table 2).

Of the 9 JBI items, 7 were satisfied. The 2 unmet criteria, absence of a control group (item 4) and lack of similar cotreatment across comparisons (item 3, driven by unadjusted concurrent dopamine or dobutamine use in over half the cohort), represent the principal sources of bias in this study and preclude a “low risk of bias” classification despite its otherwise rigorous, objectively measured, complete-follow-up design. The study is therefore judged to carry a moderate-to-high risk of bias, consistent with its role in this review as hypothesis-generating, physiological or mechanistic evidence rather than a source of causal effect estimates [26] (Table 3).

Kuragayala et al [28] scored 5 “Serious,” 1 “Moderate,” and 1 “Critical” domain on ROBINS-I, yielding an overall critical risk of bias due to the absence of any reported effect estimates, so the study is retained for descriptive purposes only (Figure 3).



XSL•FO
RenderX

**Figure 2.** Quality assessment of randomized controlled trials (RCTs) by Risk of Bias 2 (RoB 2) tool [20,21,24,25].

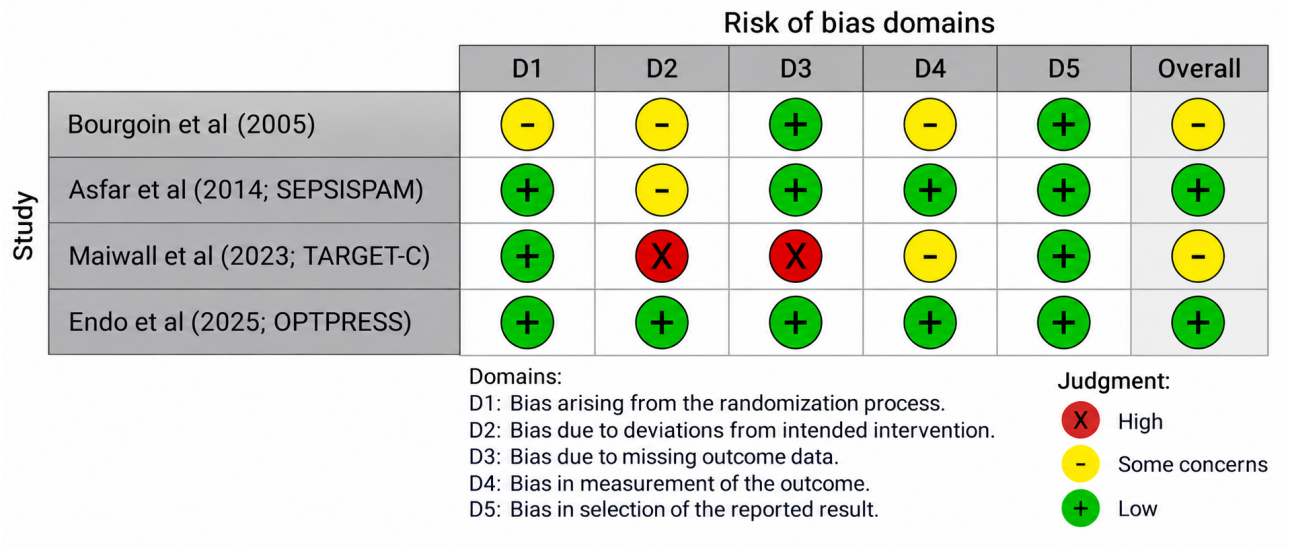



**Table .** Newcastle-Ottawa Scale—quality assessment for observational studies.

| Study | Representativeness | Selection | Exposure | Outcome absent at start | Comparability | Outcome assessment | Follow-up duration | Adequacy of follow-up | Total (/9) | Quality |
|---|---|---|---|---|---|---|---|---|---|---|
| Badin et al [22] (2011) | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 7 | Moderate |
| Poukkanen et al [23] (2013) | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 0 | 8 | Good |
| Dünser et al [27] (2009) | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 6 | Moderate |
| Dünser et al [30] (2009) | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 7 | Moderate |
| Varpula et al [29] (2005) | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 7 | Good |



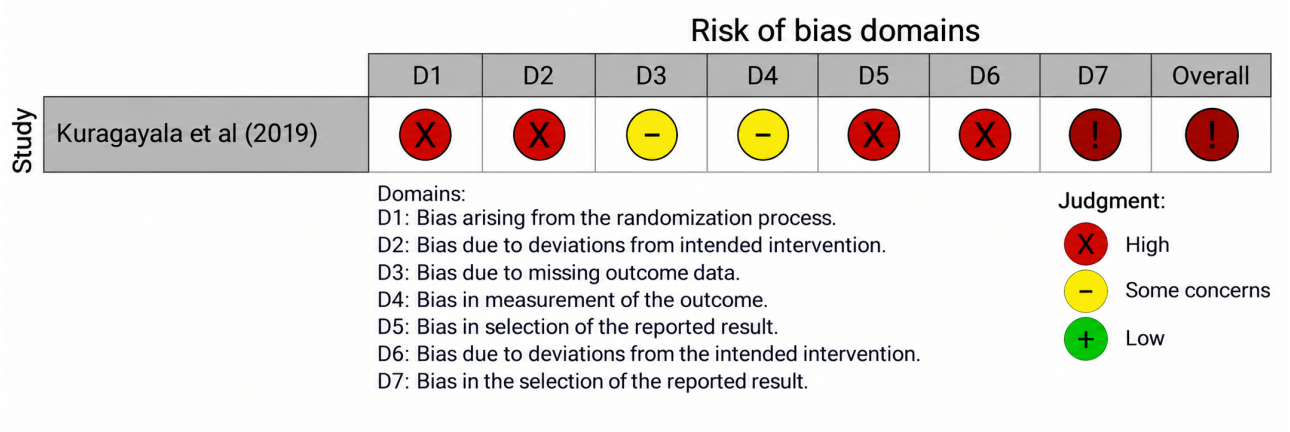
XSL•FO
RenderX

**Table .** Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Quasi-Experimental Studies.

| Item | Judgment |
|---|---|
| 1. Is it clear in the study what is the "cause" (MAP[a] titration) and what is the "effect" (perfusion or hemodynamic response)? | Yes |
| 2. Were the participants included in any comparisons similar? | Yes |
| 3. Were the participants included in any comparisons receiving similar treatment or care, other than the exposure or intervention of interest? | No |
| 4. Was there a control group? | No |
| 5. Were there multiple measurements of the outcome, both before and after the intervention or exposure? | Yes |
| 6. Was follow-up complete, and if not, were differences between groups in terms of their follow-up adequately described and analyzed? | Yes |
| 7. Were the outcomes of participants included in any comparisons measured in the same way? | Yes |
| 8. Were outcomes measured in a reliable way? | Yes |
| 9. Was appropriate statistical analysis used? | Yes |
| Overall appraisal | 7/9 "Yes" |

[a]MAP: mean arterial pressure.

**Figure 3.** Quality assessment of nonrandomized study by Risk of Bias in Non-randomized Studies of Interventions (ROBINS I) [28].




### *GRADE*

The certainty of evidence for both mortality and renal outcomes was rated as very low (Table 4). Evidence was downgraded because of methodological limitations, substantial inconsistency in the direction of treatment effects across studies, indirectness related to heterogeneous populations and MAP targets, and imprecision arising from several small studies. Consequently, the overall confidence in the effect estimates remains limited.

**Table .** Grading of Recommendations Assessment, Development, and Evaluation (GRADE) assessment.

| Outcome | Risk of bias | Inconsistency | Indirectness | Imprecision | Publication bias | Certainty |
|---|---|---|---|---|---|---|
| Mortality | Serious | Very serious | Serious | Serious | Not detected | ⊕ very low |
| AKI[a] or renal outcomes | Serious | Very serious | Serious | Serious | Not detected | ⊕ very low |

[a]AKI: acute kidney injury.

## Primary Outcome: All-Cause Mortality

### *90-Day Mortality*

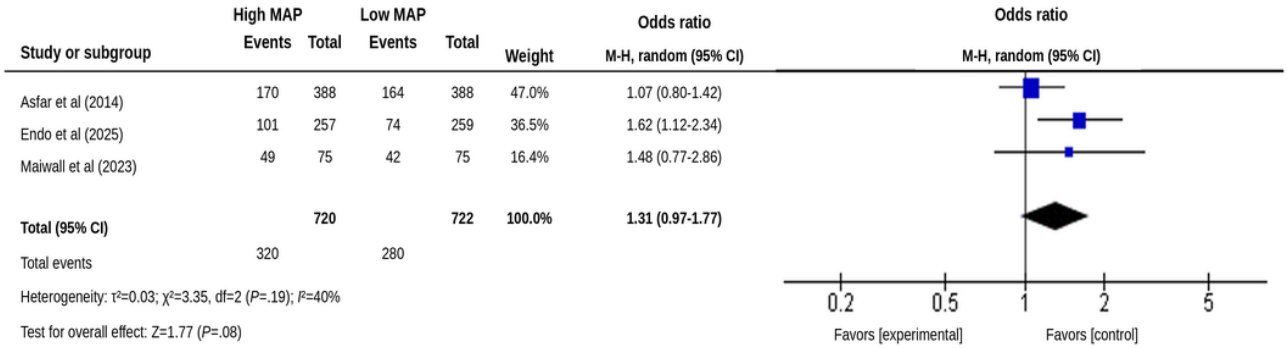
Three trials reported 90-day all-cause mortality comparing a higher versus a lower or standard MAP target and were pooled using a random-effects (Mantel-Haenszel) model [20,24,25] (Figure 4). Across 1442 patients, mortality occurred in 320 of 720 (44.4%) patients in the high-MAP arm versus 280 of 722 (38.8%) in the low-MAP arm, yielding a pooled OR of 1.31 (95% CI 0.97-1.77; $z$ score=1.77; $P$=.08), a nonsignificant trend toward higher mortality with higher MAP targets. Moderate heterogeneity was present ($\tau^2$=0.03; $\chi^2_2$=3.35, $I^2$=40%; $P$=.19). The effect was driven mainly by Endo et al [24], whose Optimal Blood Pressure in Septic Shock (OPTPRESS) trial was stopped



XSL•FO
RenderX

early for harm (OR 1.62, 95% CI 1.12-2.34) and, to a lesser extent, by Maiwall et al [20] (OR 1.48, 95% CI 0.77-2.86), while Asfar et al [25] showed an effect close to unity (OR 1.07, 95% CI 0.80-1.42; Figure 4).

**Figure 4.** Forest plot of 90-day all-cause mortality and high- versus low-mean arterial pressure (MAP) target [20,24,25].




### *28-Day Mortality and Sensitivity Analysis*

Two trials directly reported 28-day all-cause mortality by MAP-target arm [24,25]. Because Kuragayala et al [28] reported an analogous 30-day mortality comparison, a sensitivity analysis was performed with this study both added to and removed from the pooled estimate.

With Kuragayala et al [28] included (Figure 2), pooled 28- or 30-day mortality across 3 studies (695 high-MAP, 697 low-MAP patients) was 222 of 695 (32.0%) versus 191 of 697 (27.4%), giving a pooled OR of 1.25 (95% CI 0.98-1.61; $z$ score=1.78 $P$=.08), with low heterogeneity ($\chi^2_2$=2.14, $I^2$=7%; $P$=.34; Figure 5).

With Kuragayala et al [28] excluded (Figure 3), the pooled estimate from Asfar et al [25] and Endo et al [24] alone (645 patients with high-MAP, 647 patients with low-MAP) was OR of 1.31 (95% CI 0.90-1.89; $z$ score=1.43; $P$=.15), with heterogeneity rising to $I^2$ of 51% ($\chi^2_1$=2.04; $P$=.15; Figure 6).

**Figure 5.** Forest plot of 28- or 30-day mortality including Kuragayala et al [28] in the sensitivity analysis. MAP: mean arterial pressure [24,25,28].

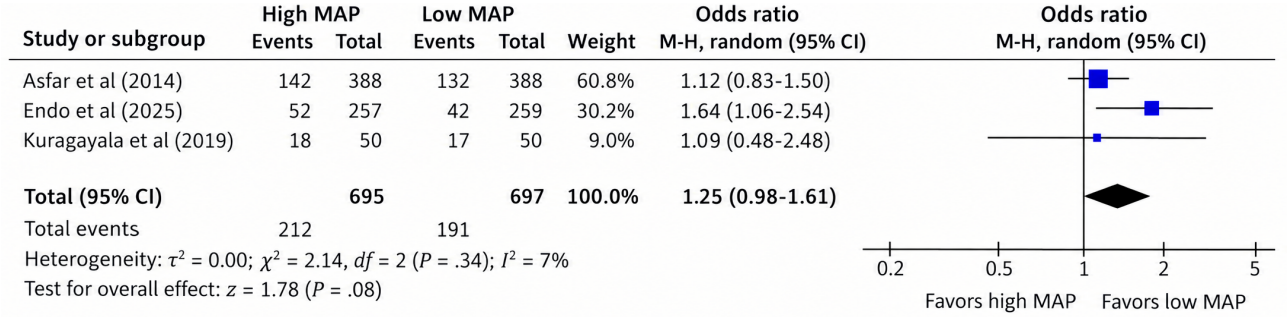
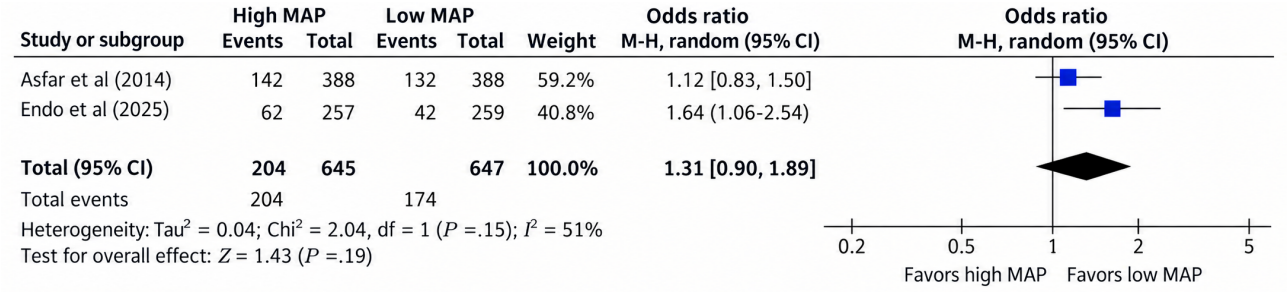



**Figure 6.** Forest plot of 28-day mortality with Kuragayala et al [28] removed in the sensitivity analysis. MAP: mean arterial pressure [24,25].




Adding Kuragayala et al [28] reduced between-study heterogeneity substantially ($I^2$ from 51% to 7%) without materially changing the point estimate or altering statistical significance, indicating the short-term mortality result was robust to this study's inclusion. Neither the 28- or 30-day nor the 90-day pooled estimate reached statistical significance, though both trended toward higher mortality with higher MAP targets.

### *Mortality Data Not Amenable to Pooling*

Several included studies reported mortality descriptively or as a regression outcome rather than as a direct between-arm comparison and could not be pooled. Bourgoin et al [21] did not assess mortality, as the trial's 8-hour observation window was designed only for physiologic and renal endpoints. Badin et al [22] reported ICU mortality of 35% and hospital mortality of 39% across the whole cohort without a formal statistical comparison by MAP exposure. In the Finnish Acute Kidney Injury Study (FINNAKI) substudy, Poukkanen et al [23] found



XSL•FO
RenderX

that 90-day mortality was significantly higher in patients whose AKI progressed (32.7% vs 18.9%; *P*=.001), though this reflected an AKI-progression subgroup rather than a randomized MAP-target comparison. In the post hoc 546C88 cohort, Dünser et al found no significant association between MAP quartile and 28-day mortality after adjustment (risk ratio [RR] 0.99, 95% CI 0.95-1.04, *P*=.82), although patients excluded for failing to sustain MAP ≥70 mm Hg had markedly higher 28-day mortality (76.5%) than the included cohort (42.1%). Conversely, Dünser et al [27] found that MAP was among the strongest independent predictors of 28-day mortality in both 6-hour (*P*=.001) and 48-hour (*P*=.01) logistic regression models, with the largest area under the receiver operating characteristic curve (AUC) for time spent below MAP 60 mm Hg (AUC=0.779) and no significant difference in discriminative performance between MAP thresholds of 60 and 75 mm Hg [30]. Similarly, Varpula et al [29] found that each 1 mm Hg decrease in mean 48-hour MAP was independently associated with an approximate 16% increase in the odds of death (Exp[B]=1.156; *P*=.01).

## Secondary Outcomes

### *AKI and RRT*

Renal outcomes were reported with heterogeneous definitions across 7 studies, precluding meta-analytic pooling. Bourgoin et al found that acutely raising MAP from 65 to 85 mm Hg over 4 hours did not change urine flow, serum creatinine, or creatinine clearance [21]. Badin et al [22] identified an MAP range of 72 to 82 mm Hg as the best discriminator of AKI progression at 72 hours, but only in patients with septic shock with early (6-h) renal impairment (AUC=0.83 - 0.84); no MAP-AKI association was found in patients without early AKI. In the larger FINNAKI substudy, Poukkanen et al [23] reported that a time-adjusted MAP below 73 mm Hg was independently associated with AKI progression after adjustment for chronic kidney disease, lactate, furosemide dose, and dobutamine use (adjusted OR 2.57, 95% CI 1.48-4.46). In Sepsis and Mean Arterial Pressure (SEPSISPAM), Asfar et al [25] found no overall between-group difference in creatinine doubling or RRT, but among patients with chronic hypertension, the high-MAP group had significantly less RRT use (31.7% vs 42.2%; *P*=.046) and less creatinine doubling (38.9% vs 52.0%; *P*=.02). In patients with cirrhosis with septic shock and AKI, Maiwall et al [20] found AKI reversal at day 5 was significantly greater with the high-MAP target in the per-protocol analysis (53% vs 31%; *P*=.02), and a high-MAP strategy independently predicted AKI recovery (OR 2.36 - 2.47). Kuragayala et al [28] reported similar, purely descriptive rates of creatinine doubling (82% vs 84%) and RRT (66% vs 70%) between arms, without inferential statistics. Finally, Dünser et al [30] found that MAP and mean perfusion pressure were both significantly correlated with RRT need ($r^2$=0.124 - 0.136; *P*≤.001) and peak creatinine ($r^2$=0.269 - 0.277), with MAP <75 mm Hg providing the best discrimination for subsequent RRT need (AUC=0.724).

### *Hemodynamic Parameters and Vasopressor Requirements*

A consistent, dose-dependent relationship between MAP target and vasopressor exposure was observed across studies. In a within-patient crossover from 65 to 75 mm Hg to 85 mm Hg, LeDoux et al [26] demonstrated a stepwise linear increase in heart rate, cardiac index, left ventricular stroke work index, systemic vascular resistance index, and norepinephrine dose (from 23 to 31 µg/min to 47 µg/min; *P*=.02). Bourgoin et al similarly found that raising MAP from 65 to 85 mm Hg significantly increased cardiac index and systemic vascular resistance index without changing lactate or oxygen consumption [21]. Among the RCTs, higher MAP targets consistently required substantially greater vasopressor exposure: Asfar et al [25] reported higher day-1 norepinephrine dose (0.58 vs 0.45 µg/kg/min; *P*<.001) and longer catecholamine infusion (4.7 vs 3.7 days; *P*<.001) in the high-MAP arm, Endo et al [24] reported longer vasopressor duration (65 vs 42 h) and higher cumulative norepinephrine (15.6 vs 9.6 mg) and vasopressin doses in the high-target arm, and Kuragayala et al [28] noted the low-MAP arm undershot its 65- to 70-mm Hg target (achieved MAP 57.3±13.7 mm Hg), while the high-MAP arm approximated its 80- to 85-mm Hg target (78.5±17.2 mm Hg).

### *Organ Support, Mechanical Ventilation, and Length of Stay*

Endo et al [24] reported significantly fewer RRT-free days (18 vs 20; *P*=.02), ventilator-free days (15 vs 18; *P*=.012), and catecholamine-free days (16 vs 19; *P*=.01) at 28 days with the high-target strategy, indicating a greater overall organ-support burden. In contrast, Asfar et al [25] found no significant difference in 28-day survival without organ support between arms (60.6% vs 62.1%; *P*=.66), and Maiwall et al [20] found no significant intention-to-treat differences in ICU length of stay or mechanical ventilation duration, although both trended longer in the high-MAP arm. In the FINNAKI substudy, patients with AKI progression had significantly longer ICU stays than those without (5.7 vs 3.8 d; *P*<.001) [23].

### *Adverse Events and Splanchnic or Ischemic Outcomes*

None of the included studies directly reported splanchnic perfusion or ischemia as a discrete primary outcome; the closest available proxy was mesenteric ischemia as an adverse event, which occurred at similar, low rates between arms in both Asfar et al [25] (2.3% vs 2.3%) and Maiwall et al [20] (1.3% high-MAP vs 0% low-MAP). Cardiac arrhythmias were more frequent with higher MAP targets in several studies: Asfar et al [25] reported significantly more atrial fibrillation with high MAP (6.7% vs 2.8%, *P*=.02), and Endo et al [24] noted that all main safety outcomes—arrhythmia, ischemic events, and hemorrhagic events—trended more frequent in the high-target group, although none reached statistical significance individually. Maiwall et al [20] found significantly more adverse events requiring protocol discontinuation with high MAP (24% vs 11%; *P*=.03), but significantly less intradialytic hypotension (12% vs 48%; *P*<.001), suggesting a trade-off between systemic vasopressor-related harm and protection from dialysis-associated hypotension at higher MAP targets.



XSL•FO
RenderX

## Discussion

### Principal Findings

This systematic review synthesized evidence from 11 studies examining MAP targets during septic shock resuscitation, with particular attention to splanchnic perfusion and renal outcomes. Pooled analysis of 3 trials [20,24,25] showed a nonsignificant trend toward higher 90-day mortality with higher MAP targets (OR 1.31, 95% CI 0.97-1.77), a trend echoed in the pooled 28- or 30-day mortality estimate derived from Asfar et al [25], Endo et al [24], and, in sensitivity analysis, Kuragayala et al [28] (OR 1.25, 95% CI 0.98-1.61). Although neither estimate reached statistical significance, the consistency of direction across both time points, combined with the early termination of the OPTPRESS trial due to harm [24], raises genuine concern that targeting a higher MAP does not confer a survival benefit and may plausibly increase risk in some patients.

These findings diverge from the physiological rationale that historically supported higher MAP targets in septic shock. Early mechanistic work by LeDoux et al [26] and Bourgoin et al [21] showed that acutely raising MAP from 65 to 85 mm Hg increased cardiac index and systemic vascular resistance without improving lactate clearance, urine flow, or creatinine clearance, suggesting that macrocirculatory augmentation above 65 mm Hg does not reliably translate into organ-level perfusion benefit. This disconnect persisted into the RCT era: Asfar et al [25] found no overall mortality or organ support–free survival benefit with an 80 - to 85-mm Hg target, and Endo et al [11] found the higher target reduced RRT-free, ventilator-free, and catecholamine-free days at 28 days.

A recurring theme across the evidence base, however, is that the relationship between MAP and organ injury is not uniform across patients and that baseline physiology modifies both the direction and magnitude of treatment effect. The clearest example is chronic hypertension: within SEPSISPAM, Asfar et al [25] reported that patients with hypertension randomized to the higher MAP target had significantly less RRT use (31.7% vs 42.2%; *P*=.046) and less creatinine doubling (38.9% vs 52.0%; *P*=.02), an effect not observed in patients with normal blood pressure. This is consistent with a rightward-shifted renal autoregulatory curve in patients with chronic hypertension, whereby an MAP adequate for autoregulation in normotensive individuals may fall below the autoregulatory threshold in those with long-standing hypertension. A second example is cirrhosis: Maiwall et al [20] found that a higher MAP target improved AKI reversal at day 5 in patients with cirrhosis and septic shock and AKI (53% vs 31%; *P*=.02), a population characterized by splanchnic vasodilation, effective arterial underfilling, and impaired renal autoregulation distinct from the general septic shock population. These subgroup signals, alongside the broader trial-level null or harm findings, suggest that a single population-wide MAP target may obscure clinically important heterogeneity of treatment effect and that individualized, comorbidity-informed targets may be more physiologically appropriate than fixed thresholds.

The renal outcome data reinforce this picture of MAP as a threshold rather than a linear "higher is better" variable. Poukkanen et al [23] identified a time-adjusted MAP below 73 mm Hg as independently associated with AKI progression (adjusted OR 2.57, 95% CI 1.48-4.46), while Dünser et al [30] reported that time below an MAP of 60 mm Hg was the strongest discriminator of 28-day mortality (AUC=0.779), with no added benefit from thresholds between 60 and 75 mm Hg. Similarly, Badin et al [22] found that an MAP of 72 - 82 mm Hg best discriminated AKI progression only in patients with existing early renal impairment. Collectively, these data suggest a threshold effect in which avoiding hypotension below roughly 60 to 65 mm Hg is critical, but pushing MAP higher confers diminishing or absent renal benefit, consistent with the null or harmful mortality signals from Asfar et al [25] and Endo et al [24].

Notably, none of the included studies reported splanchnic perfusion or ischemia as a discrete primary outcome, despite this being a central rationale for the review. The only available proxy, mesenteric ischemia as an adverse event, occurred at low and similar rates between MAP arms in both Asfar et al [25] (2.3% vs 2.3%) and Maiwall et al [20] (1.3% vs 0%), providing some reassurance that higher MAP targets are not associated with excess overt bowel ischemia, but this outcome was underpowered and nonspecific and cannot be equated with subclinical splanchnic hypoperfusion [20,25]. This represents an important evidence gap rather than evidence of absence of effect.

The safety profile of higher MAP targets also warrants attention. Higher targets were consistently associated with greater vasopressor exposure, from the stepwise norepinephrine increases seen by LeDoux et al to the significantly higher day-1 and cumulative catecholamine doses reported by Asfar et al [25] and Endo et al [24,26]. This increased vasopressor burden was accompanied by a higher incidence of atrial fibrillation in Asfar et al [25] (6.7% vs 2.8%; *P*=.02) and a trend toward more arrhythmic, ischemic, and hemorrhagic events in Endo et al [24]. Maiwall et al [20] identified a clinically important trade-off in patients with cirrhosis, in whom higher MAP targets increased protocol-discontinuing adverse events (24% vs 11%; *P*=.03) but substantially reduced intradialytic hypotension (12% vs 48%; *P*<.001), illustrating that the net safety balance of a given MAP target may depend heavily on which complications are most relevant to a specific population.

Our findings both extend and refine the conclusions of prior reviews. Leone et al [8], synthesizing studies through July 2014, concluded that an MAP of 65 mm Hg was generally sufficient, although 75 to 85 mm Hg might reduce AKI in patients with chronic hypertension. Our review, incorporating a decade of additional evidence including SEPSISPAM, Targeted Mean Arterial Pressure in Cirrhosis with Septic Shock (TARGET-C), and OPTPRESS, corroborates this hypertension-specific renal signal but adds a caveat Leone et al [8] could not have anticipated: the subsequent OPTPRESS trial [24], conducted in older patients, was stopped early for harm, indicating that higher-MAP benefit in hypertensive or older populations is not generalizable and may reverse in advanced age. Dari et al [10], restricting their meta-analysis to 4 RCTs, similarly found no significant difference in mortality (RR 0.94, 95% CI 0.87-1.01), RRT use, or ICU stay between high- and low-normal MAP



XSL•FO
RenderX

targets. However, by excluding observational evidence entirely, Dari et al [10] could not capture the continuous, threshold-based MAP–outcome relationship shown in cohorts such as Poukkanen et al [23], Varpula et al [29], and Dünser et al [27,30], nor the cirrhosis-specific findings of Maiwall et al [20] that our broader inclusion criteria captured. Kattan et al [11], reviewing hemodynamic monitoring, argued for shifting away from fixed macrohemodynamic targets such as MAP toward multimodal, perfusion-based endpoints such as capillary refill time. Our findings lend indirect support: the absence of consistent mortality or renal benefit from higher MAP across heterogeneous populations, alongside dose-dependent escalation in vasopressor exposure and arrhythmic risk, suggests that a single macrohemodynamic number is an imprecise proxy for tissue perfusion, reinforcing the case for perfusion-guided, individualized resuscitation over uniform MAP thresholds.

### Limitations

Several limitations should be considered. First, substantial clinical and methodological heterogeneity across included studies, spanning populations (general septic shock, cirrhosis, older adults-only cohorts), MAP comparison strategies (fixed targets vs continuous exposure), and outcome definitions, precluded meta-analysis for all outcomes except mortality. Second, the overall certainty of evidence for mortality and renal outcomes was rated as very low using GRADE, reflecting serious risk of bias, inconsistency, and imprecision. Third, several trials were open-label, and 2 RCTs [20,21] were rated as having some concerns on RoB 2. Fourth, this review was not prospectively registered on PROSPERO, and restriction to English-language, full-text-available studies may have introduced language bias. Fifth, reliance on Google Scholar screening limited to the first 400 relevance-ranked results may have missed additional eligible studies. Sixth, this review itself included only 4 RCTs, a limitation similar to those it critiques in prior reviews. Seventh, adverse events were systematically reported in only 3 of the included studies, and arrhythmia data derive from just 2 trials (Asfar et al [25], Endo et al [24]), limiting the generalizability of the safety findings. Finally, no included study directly measured splanchnic perfusion or ischemia, forcing reliance on an imprecise adverse-event proxy that cannot capture subclinical gut hypoperfusion.

### Future Directions

Future research should prioritize direct, quantitative measurement of splanchnic and mesenteric perfusion, using gastric tonometry, sublingual microcirculatory imaging, near-infrared spectroscopy, or gut-barrier biomarkers, within randomized MAP-target comparisons. Given the signal of differential treatment effect by comorbidity, adequately powered trials or individual patient data meta-analyses stratified by chronic hypertension, cirrhosis, and age are needed to clarify which subgroups benefit from higher MAP targets, and to reconcile the hypertension-related benefit reported by Leone et al [8] and Asfar et al [25] with the harm signal in older patients seen in Endo et al [24]. Trials should incorporate standardized vasopressor and adverse-event reporting to enable pooled safety analyses and should consider individualized or perfusion-guided targets, as advocated by Kattan et al [11], rather than fixed population thresholds. Finally, prospective registration and consistent outcome definitions across future studies would improve the poolability of evidence in this field.

### Conclusion

Overall, this review does not support routinely targeting a higher MAP in septic shock, as the available evidence points toward no meaningful survival benefit and, if anything, a possible increase in harm, alongside greater vasopressor exposure and cardiovascular risk. At the same time, the evidence suggests that MAP is not a truly one-size-fits-all target: certain patients, particularly those with chronic hypertension or underlying liver disease, may benefit from a somewhat higher pressure target to protect kidney function, reflecting differences in how these patients' bodies regulate blood flow. Importantly, none of the available studies directly examined gut perfusion, leaving the central question motivating this review unanswered. Given the overall low certainty of the current evidence, a standard MAP target remains the safest general approach for most patients with septic shock, while an individualized, slightly higher target may be reasonable in select high-risk subgroups. Future research should move beyond a single blood pressure number and instead focus on directly assessing tissue perfusion, tailoring resuscitation strategies to the individual patient rather than applying a uniform target to all.

## Acknowledgments

An AI tool (Claude Opus 5.5) was used in data extraction, but statistics were cross-checked for all extractable characteristics and results of the studies. It was also used in making tables.

## Funding

The authors declare no financial support was received for this work.

## Authors' Contributions

Conceptualization: PC (lead), JFOC (equal), LMPL (supporting)
Data curation: NQ (lead), VGOC (supporting)
Formal analysis: PC (lead), JFOC (equal), NQ (supporting)
Funding acquisition: PC
Investigation: LMPL (lead), JDUO (supporting), EPCA (supporting)



XSL•FO
RenderX

Methodology: PC (lead), LMPL (equal), JFOC (supporting)
Project administration: PC (lead), JFOC (equal), VGOC (supporting)
Resources: LMPL (lead), EPCA (supporting)
Supervision: PC (lead), JFOC (supporting)
Validation: PC (lead), LMPL (supporting)
Visualization: VGOC (lead), NQ (supporting)
Writing – original draft: PC (lead), JFOC (equal), LMPL (supporting), NQ (supporting)
Writing – review & editing: PC (lead), JFOC (equal), LMPL (supporting), NQ (supporting), VGOC (supporting), JDUO (supporting), EPCA (supporting)

## Conflicts of Interest

None declared.

Checklist 1
PRISMA checklist.
[PDF File, 180 KB - i-jmr_v15i1e98772_app1.pdf ]

## References


1. Rudd KE, Johnson SC, Agesa KM, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. Lancet 2020 Jan 18;395(10219):200-211. [doi: 10.1016/S0140-6736(19)32989-7] [Medline: 31954465]
2. GBD 2021 Global Sepsis Collaborators. Global, regional, and national sepsis incidence and mortality, 1990-2021: a systematic analysis. Lancet Glob Health 2025 Dec;13(12):e2013-e2026. [doi: 10.1016/S2214-109X(25)00356-0] [Medline: 41135560]
3. Sands K, Carvalho MJ, Portal E, et al. Characterization of antimicrobial-resistant Gram-negative bacteria that cause neonatal sepsis in seven low- and middle-income countries. Nat Microbiol 2021 Apr;6(4):512-523. [doi: 10.1038/s41564-021-00870-7] [Medline: 33782558]
4. Evans L, Rhodes A, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Med 2021 Nov;47(11):1181-1247. [doi: 10.1007/s00134-021-06506-y] [Medline: 34599691]
5. Moman RN, Ostby SA, Akhoundi A, Kashyap R, Kashani K. Impact of individualized target mean arterial pressure for septic shock resuscitation on the incidence of acute kidney injury: a retrospective cohort study. Ann Intensive Care 2018 Dec 10;8(1):124. [doi: 10.1186/s13613-018-0468-5] [Medline: 30535664]
6. Xu JY, Ma SQ, Pan C, et al. A high mean arterial pressure target is associated with improved microcirculation in septic shock patients with previous hypertension: a prospective open label study. Crit Care 2015 Mar 30;19(1):130. [doi: 10.1186/s13054-015-0866-0] [Medline: 25887027]
7. García-de-Acilu M, Mesquida J, Gruartmoner G, Ferrer R. Hemodynamic support in septic shock. Curr Opin Anaesthesiol 2021 Apr 1;34(2):99-106. [doi: 10.1097/ACO.0000000000000959] [Medline: 33652455]
8. Leone M, Asfar P, Radermacher P, Vincent JL, Martin C. Optimizing mean arterial pressure in septic shock: a critical reappraisal of the literature. Crit Care 2015;19(1):101. [doi: 10.1186/s13054-015-0794-z] [Medline: 25888071]
9. Kędziora A, Piątek J, Hymczak H, et al. Early postoperative hemodynamic instability after heart transplantation—incidence and metabolic indicators. BMC Anesthesiol 2021 Oct 2;21(1):236. [doi: 10.1186/s12871-021-01455-x] [Medline: 34600476]
10. Dari MA, Fayaz A, Sharif S, et al. Comparison of high-normal versus low-normal mean arterial pressure at target on outcomes in sepsis or shock patients: a meta-analysis of randomized control trials. Cureus 2024 Jan;16(1):e52258. [doi: 10.7759/cureus.52258] [Medline: 38352092]
11. Kattan E, Castro R, Vera M, Hernández G. Optimal target in septic shock resuscitation. Ann Transl Med 2020 Jun;8(12):789. [doi: 10.21037/atm-20-1120] [Medline: 32647714]
12. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
13. Pettilä V, Pfortmüller CA, Perner A, et al. Targeted tissue perfusion versus macrocirculatory-guided standard care in patients with septic shock: a randomized clinical trial-the TARTARE-2S trial. Crit Care Med 2026 Jan 1;54(1):24-34. [doi: 10.1097/CCM.0000000000006899] [Medline: 41105050]
14. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
15. Gierisch JM, Beadles C, Shapiro A, et al. Appendix b—newcastle-ottawa scale coding manual for cohort studies. In: Health Disparities in Quality Indicators of Healthcare Among Adults with Mental Illness: Department of Veterans Affairs (US); 2014. URL: https://www.ncbi.nlm.nih.gov/books/NBK299087 [accessed 2026-09-01]




XSL•FO
RenderX

16. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919. [doi: 10.1136/bmj.i4919] [Medline: 27733354]
17. Barker TH, Habibi N, Aromataris E, et al. The revised JBI critical appraisal tool for the assessment of risk of bias for quasi-experimental studies. JBI Evid Synth 2024 Mar 1;22(3):378-388. [doi: 10.11124/JBIES-23-00268] [Medline: 38287725]
18. Campbell M, McKenzie JE, Sowden A, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ 2020 Jan 16;368:l6890. [doi: 10.1136/bmj.l6890] [Medline: 31948937]
19. Prasad M. Introduction to the GRADE tool for rating certainty in evidence and recommendations. Clin Epidemiol Glob Health 2024 Jan;25:101484. [doi: 10.1016/j.cegh.2023.101484]
20. Maiwall R, Rao Pasupuleti SS, Hidam AK, et al. A randomised-controlled trial (TARGET-C) of high vs. low target mean arterial pressure in patients with cirrhosis and septic shock. J Hepatol 2023 Aug;79(2):349-361. [doi: 10.1016/j.jhep.2023.04.006] [Medline: 37088310]
21. Bourgoin A, Leone M, Delmas A, Garnier F, Albanèse J, Martin C. Increasing mean arterial pressure in patients with septic shock: effects on oxygen variables and renal function. Crit Care Med 2005 Apr;33(4):780-786. [doi: 10.1097/01.ccm.0000157788.20591.23] [Medline: 15818105]
22. Badin J, Boulain T, Ehrmann S, et al. Relation between mean arterial pressure and renal function in the early phase of shock: a prospective, explorative cohort study. Crit Care 2011;15(3):R135. [doi: 10.1186/cc10253] [Medline: 21645384]
23. Poukkanen M, Wilkman E, Vaara ST, et al. Hemodynamic variables and progression of acute kidney injury in critically ill patients with severe sepsis: data from the prospective observational FINNAKI study. Crit Care 2013 Dec 13;17(6):R295. [doi: 10.1186/cc13161] [Medline: 24330815]
24. Endo A, Yamakawa K, Tagami T, et al. Efficacy of targeting high mean arterial pressure for older patients with septic shock (OPTPRESS): a multicentre, pragmatic, open-label, randomised controlled trial. Intensive Care Med 2025 May;51(5):883-892. [doi: 10.1007/s00134-025-07910-4] [Medline: 40358717]
25. Asfar P, Meziani F, Hamel JF, et al. High versus low blood-pressure target in patients with septic shock. N Engl J Med 2014 Apr 24;370(17):1583-1593. [doi: 10.1056/NEJMoa1312173] [Medline: 24635770]
26. LeDoux D, Astiz ME, Carpati CM, Rackow EC. Effects of perfusion pressure on tissue perfusion in septic shock. Crit Care Med 2000 Aug;28(8):2729-2732. [doi: 10.1097/00003246-200008000-00007] [Medline: 10966242]
27. Dünser MW, Ruokonen E, Pettilä V, et al. Association of arterial blood pressure and vasopressor load with septic shock mortality: a post hoc analysis of a multicenter trial. Crit Care 2009;13(6):R181. [doi: 10.1186/cc8167] [Medline: 19917106]
28. Kuragayala SD, Ganti SR, Sirga S, Kommula HB. Ideal mean arterial pressure target in septic shock: the hunt goes on!. Saudi J Med 2019;4:664-667. [doi: 10.36348/sjm.2019.v04i08.021]
29. Varpula M, Tallgren M, Saukkonen K, Voipio-Pulkki LM, Pettilä V. Hemodynamic variables related to outcome in septic shock. Intensive Care Med 2005 Aug;31(8):1066-1071. [doi: 10.1007/s00134-005-2688-z] [Medline: 15973520]
30. Dünser MW, Takala J, Ulmer H, et al. Arterial blood pressure during early sepsis and outcome. Intensive Care Med 2009 Jul;35(7):1225-1233. [doi: 10.1007/s00134-009-1427-2] [Medline: 19189077]

## Abbreviations

**AKI:** acute kidney injury
**APACHE II:** Acute Physiology and Chronic Health Evaluation II
**AUC:** area under the receiver operating characteristic curve
**CENTRAL:** Cochrane Central Register of Controlled Trials
**FINNAKI:** Finnish Acute Kidney Injury Study
**GRADE:** Grading of Recommendations Assessment, Development, and Evaluation
**ICU:** intensive care unit
**JBI:** Joanna Briggs Institute
**MAP:** mean arterial pressure
**NOS:** Newcastle-Ottawa Scale
**OPTPRESS:** Optimal Blood Pressure in Septic Shock
**OR:** odds ratio
**PICO:** Population, Intervention, Comparator, and Outcome
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**RCT:** randomized controlled trial
**RoB 2:** Risk of Bias 2
**ROBINS-I:** Risk of Bias in Non-randomized Studies of Interventions
**RR:** risk ratio
**RRT:** renal replacement therapy
**SAPS II:** Simplified Acute Physiology Score II
**SEPSISPAM:** Sepsis and Mean Arterial Pressure
**SOFA:** Sequential Organ Failure Assessment



XSL•FO
RenderX

**TARGET-C:** Targeted Mean Arterial Pressure in Cirrhosis with Septic Shock

*Edited by M Balcarras; submitted 18.Apr.2026; peer-reviewed by C Ngan, W Tambo; revised version received 19.Aug.2026; accepted 20.Aug.2026; published 27.Oct.2026.*

*Please cite as:*
*Cisneros P, Ortiz Caviedes JF, Piñeros Ladino LM, Quintero N, Oñate Cervera VG, Urrea Ospina JD, Cerna Alcantar EP*
*Impact of Elevated Blood Pressure Targets on Splanchnic Ischemia and Mortality in Septic Resuscitation: Systematic Review*
*Interact J Med Res 2026;15:e98772*
*URL: https://www.i-jmr.org/2026/1/e98772*
*doi:10.2196/98772*

© Paulina Cisneros, John Fredy Ortiz Caviedes, Lina Maria Piñeros Ladino, Natalia Quintero, Valeria García Oñate Cervera, Juan David Urrea Ospina, Erika Patricia Cerna Alcantar. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 27.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Overcrowding Indicators in Emergency Departments Across Countries: Scoping Review

Natasya Nur Mohd Nasir*, MD; Ku Anis Shazura Indera Putera*, MSc; Nur Jihan Noris*, MD; Zalina Libasin*, MBBS; Muniamal Krishnan*, MD, MOSHRM; Nor Fauziah Salaton*, DrPH; Kah Yee Lum*, MSc; Nur Nadia Renu Abdullah*, MNSc; Intan Syafinaz Saimy*, DrPH

Institute for Health Management, National Institutes of Health, Block B1, No.1, Jalan Setia Murni U13/52, Section U13, Shah Alam, Selangor, Malaysia
*all authors contributed equally

**Corresponding Author:**
Natasya Nur Mohd Nasir, MD
Institute for Health Management, National Institutes of Health, Block B1, No.1, Jalan Setia Murni U13/52, Section U13, Shah Alam, Selangor, Malaysia

## Abstract

**Background:** Emergency department (ED) overcrowding is a persistent global health issue associated with adverse patient outcomes, diminished staff performance, and compromised health-system efficiency. Despite widespread recognition of the problem, there is no universally accepted approach to monitoring ED overcrowding. The use of disparate, nonstandardized indicators hampers cross-country comparison and the development of effective policies. A comprehensive synthesis of indicators currently used is essential to guide the adoption of robust, evidence-based metrics across diverse health care settings.

**Objective:** This study aims to identify, consolidate, and categorize indicators that have been used internationally to assess ED overcrowding and to highlight gaps in their use.

**Methods:** A comprehensive scoping review was conducted from October to November 2023 using four databases: PubMed, Scopus, Emerald Insight, and Google Scholar. Studies were systematically searched using predefined eligibility criteria. Level 1 and 2 screening were independently conducted by 9 researchers (NNMN, KASIP, NFS, NJN, MK, ZL, NNRA, LKY, and ISS) to minimize bias and enhance reliability, and discrepancies were resolved by consensus. A third reviewer (ISS) performed a full-text review, synthesis, and descriptive analysis. Indicators were categorized into input, throughput, and output. Input refers to factors driving ED demand, throughput encompasses internal ED processes such as triage, diagnostics, and treatment, and output addresses challenges in transferring patients to inpatient beds, such as bed shortages or delays. Descriptive analyses were then used to consolidate these indicators and to establish their relative importance. They were ranked based on frequency of reporting across diverse countries and health care settings.

**Results:** Out of 1347 articles screened, 117 articles were included in the study. A total of 307 indicators were retrieved and then consolidated into 26 distinct indicators. The majority of indicators were classified within the throughput domain (209/307, 68%), followed by the output domain (62/307, 20%) and the input domain (36/307, 12%). The most common throughput indicator, which was frequently reported, was ED length of stay, cited 87 times, followed by left without being seen and waiting time, each reported 30 times. Length of stay consistently emerged as a primary marker of systemic bottlenecks and operational inefficiencies across health care systems.

**Conclusions:** This review indicates that throughput measures, particularly length of stay, dominate current approaches to assessing ED overcrowding, whereas input and output indicators remain comparatively underrepresented. By consolidating 26 distinct indicators from 117 studies, this study provides a comprehensive evidence base to support the standardization of metrics for monitoring ED overcrowding internationally. These findings offer practical guidance for policymakers and health care leaders seeking to refine performance indicators, enhance benchmarking, and evaluate interventions aimed at improving patient flow. Further research should prioritize validation of underused indicators and the development of composite measures that better capture the complexity of ED crowding across diverse health care settings.

***(Interact J Med Res 2026;15:e78073)*** doi:10.2196/78073

**KEYWORDS**

emergency service; emergency indicator; emergency department; overcrowding indicators; emergency department overcrowding; global emergency overcrowding; PRISMA



XSL•FO
RenderX

## Introduction

### Background

Emergency department (ED) overcrowding is a significant challenge faced by health care systems globally [1]. Although various studies have proposed definitions, ED overcrowding fundamentally occurs when the department's capacity is insufficient to meet the urgent needs of patients, leading to delays in care due to congestion [2]. Overcrowding is characterized by a surge in patients seeking care that exceeds the department's capacity to provide timely and efficient services, which reflects a mismatch between patient demand and available resources [3]. Overcrowding in the ED can arise due to a variety of factors, including an increase in patient volume, staffing shortages, and inefficient workflows. The scenario of overcrowding often begins even before a patient sets foot in the ED, influenced by factors such as external medical systems, lack of timely access to primary care, or seasonal health surges [4].

It is also associated with a range of adverse outcomes, including longer treatment durations, preventable medical errors, and the proportion of patients who leave the ED without receiving a medical evaluation from a health care professional [5]. Addressing ED overcrowding effectively requires comprehensive monitoring of patient flow, guided by a well-structured system of performance indicators. Health care authorities rely on these quality indicators to assess ED performance during regular operations and identify opportunities for improvement. The selection of appropriate indicators is crucial, as they guide decision-making based on specific process factors and the outcomes they measure. Improper selection or implementation of these indicators can lead to unintended and detrimental effects [6].

Indicators are essential for addressing ED overcrowding by providing real-time, data-driven insights across the patient journey, from seeking care to discharge or admission, while also monitoring key quality metrics to ensure efficient care delivery [7]. Both operational and clinical indicators are vital for guiding decision-making at each level. Previsit indicators, such as emergency call volume trends, ambulance diversion rates, and regional health data, help predict demand surges. Once patients arrive, operational indicators like triage times, treatment delays, and bed availability ensure efficient resource allocation and timely care delivery. Monitoring these indicators is crucial for public health surveillance, as they provide a comprehensive understanding of overcrowding and inform evidence-based interventions [8]. By tracking specific indicators, health care authorities can identify emerging trends, proactively allocate resources, and implement strategies to prevent overcrowding, thereby improving patient outcomes and health care delivery. Similarly, an Australian ED is exploring the integration of patient flow indicators, including measures of patient acuity, length of stay, and the availability of beds, to improve patient care and manage overcrowding in the ED. The implementation of the "Emergency Department Throughout Time Indicator" in Australia has helped to streamline patient flow, reducing waiting times for critical care [9].

The need for appropriate ED overcrowding indicators is critical for enhancing the efficiency and effectiveness of emergency care systems [10]. By using these indicators, health care authorities can analyze key metrics, such as patient wait times, length of stay, and throughput rates, providing a more accurate understanding of patient flow dynamics and facilitating targeted interventions [11]. These indicators play a vital role in assessing resource use and availability, enhancing the understanding of overcrowding, and driving improvements in health care delivery, resource management, and patient outcomes in both global and local contexts [6].

### Study Objective and Rationale

The primary objective of this study is to describe the indicators used to measure overcrowding at the EDs in hospitals using a scoping review. This objective helps to review the indicator used across a range of hospital settings globally. To achieve this, the study is guided by the central research question: what indicators are used to measure overcrowding in EDs in hospitals worldwide? This question forms the basis of the investigation, which aims to provide a comprehensive understanding of how overcrowding is measured and managed across different health care environments. The findings are expected to contribute to the development of more effective strategies for addressing ED overcrowding and improving health care delivery on a global scale.

## Methods

### Study Design

This scoping review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Review and Meta-Analyses extension for Scoping Reviews) as presented in Checklist 1.

### Protocol and Registration

The study protocol was registered with the National Medical Research Registry, Malaysia, under NMRR ID-23 - 01709-HDO. Data collection was carried out between May 2023 and April 2024. The content of this review adheres to the registered protocol and aligns with the stated study objectives. The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

### Eligibility Criteria

The search strategy was developed using the population, intervention, and outcome framework. The study population comprised health institutions with operational EDs, where service congestion and overcrowding are frequently reported. The intervention was defined as the implementation, application, or evaluation of institutional or hospital performance measures. The outcome focused on the primary indicators used to assess ED overcrowding within these emergency care settings. The comprehensive set of keywords related to ED overcrowding and the indicators used to evaluate or manage it was used to identify relevant studies. All records retrieved from the initial



XSL•FO
RenderX

search were screened and assessed for relevance according to predefined inclusion and exclusion criteria in the Textbox 1.

**Textbox 1.** Eligibility criteria.

**Inclusion Criteria**

Studies were included if they met the following criteria:

- **Population:** conducted in hospitals, health centers, health institutions, emergency departments (EDs), trauma centers, or accident and EDs.
- **Intervention:** investigated indicators, measures, indices, parameters, attributes, determinants, or metric indices related to ED performance.
- **Outcome:** reported outcomes associated with overcrowding, congestion, or service overload within ED settings.
- **Study design:** primary research studies of any design, including reports addressing ED overcrowding.
- **Publication type:** article published in peer-reviewed journals.
- **Language:** published in English.
- **Access:** full-text articles available for review.

**Exclusion Criteria**

Studies were excluded if they met any of the following criteria:

- **Population:** not focused on EDs or involved nonhuman populations or nonhealth care facilities (eg, veterinary hospitals).
- **Intervention:** did not include any indicators or measurements related to ED overcrowding.
- **Outcome:** did not address issues of overcrowding, congestion, or overload.
- **Study design:** nonprimary studies such as literature reviews, commentaries, study protocols, or ongoing research.
- **Publication type:** other publication types, including conference papers, dissertations, or abstract books.
- **Language:** published in languages other than English.
- **Access:** full-text article not available.

## Information Sources

Articles were sourced from selected databases, including PubMed, Emerald Insight, Google Scholar, and Scopus. Google Scholar was included to capture gray literature, as the initial search yielded a substantial number of documents, particularly theses and technical reports. All retrieved records were assessed for eligibility.

## Search Strategy

The search process was conducted by the principal investigator, KASIP, in June 2023. This process targeted studies published between 2013 and 2023 to ensure the inclusion of the most recent evidence, considering health care system advancements [12]. The inclusion timeframe was restricted to the past 10 years to capture the most recent evidence, considering the rapidly evolving health care landscape that may limit the applicability of older studies [13]. A research specialist reviewed and optimized the search strategy, focusing on Boolean operators and database-specific indexing terms. Keywords were combined into final search strings, such as (hospital OR health OR healthcare) AND (indicators OR measure OR index) AND (overcrowding OR congestion), which were applied across all databases. The number of records retrieved and the corresponding search strategies are summarized in Table S1 in Multimedia Appendix 1. All results were exported to Mendeley, and duplicates were removed manually. The final set of articles meeting the inclusion criteria was incorporated into the review.

## Study Selection

Selected studies that capture literature on ED overcrowding and relevant indicators based on the eligibility criteria Textbox 1 were managed using Mendeley software by KASIP and LKY. The studies were evenly distributed among 4 pairs of researchers (Group 1: NJN and NNRA, Group 2: NNMN and ISS, Group 3: ZL and LKY, and Group 4: KASIP, NFS, and MK). Nine researchers (NNMN, KASIP, NFS, NJN, MK, ZL, NNRA, LKY, and ISS), working in pairs, independently reviewed the same articles to ensure accuracy and consistency. In most cases, each pair included at least 1 researcher with a clinical or public health background. Specifically, KASIP is a research officer, NNRA is a nurse researcher, and LKY is a pharmacist. During the level-1 screening process, researchers reviewed the titles and abstracts of the studies, and the required information was recorded using the Google Form. In level-2 screening, the process involved a detailed evaluation of the full texts of the studies that passed the first stage, with results recorded in a Google Sheet. Disagreements between paired reviewers at both stages were initially resolved through discussion and consensus. If the consensus was not achieved, a third independent reviewer (ISS) from a different screening pair, who had not previously screened the study, acted as the assessor to make the final decision.

## Data Charting

During level-1 screening, data collected primarily included study characteristics such as focus on ED overcrowding, document type, study design, publication status, and the presence of reported indicators. Upon completion of this stage, findings



XSL•FO
RenderX

were mapped based on the presence of indicators or other measurements, and studies reporting relevant data were selected for level-2 screening. Level-2 screening focused on more detailed information, including year of publication, authorship, study setting, study design, study objectives, study acceptance, identified indicators, and the specific measurement tools used by 4 pairs. Findings from both screening levels were cross-checked, and any discrepancies were resolved through consensus discussions between reviewers to ensure rigor, validity, and trustworthiness. KASIP made the final determination, which was subsequently validated by ISS.

### Data Items

A list of data items was developed by KASIP and subsequently validated by ISS. These items are presented in Table 1.

**Table .** Data items extracted at each screening level.

| Level of screening | Data items |
|---|---|
| Level 1 | • Focus on ED[a] overcrowding (yes, no, or unable to retrieve the abstract)<br>• Document type (eg, articles in a journal: review, letters, editorial or comments, consensus statement, newsletter, original article, policy brief, thesis, or report)<br>• Study design (eg, case control, case study, cohort, cross-sectional, mixed methods, qualitative study, quasi-experiment study, randomized controlled trial, reviews – scoping or systematic and literature)<br>• Publication status - ongoing studies or protocols (yes or no)<br>• The presence of reported indicator (yes or not applicable) |
| Level 2 | • Article ID (code assigned to each selected study)<br>• Year of publication<br>• List of authors<br>• Study setting (origin or site of the study implementation)<br>• Study design<br>• Study objectives<br>• Acceptance of the study (accept or reject)<br>• Identified indicators<br>• Specific measurement tools (eg, National ED Overcrowding Study (NEDOCS) and Emergency Department Work Index (EDWIN) |

[a]ED: emergency department.

### Critical Appraisal

Critical appraisal was not conducted and no risk of bias assessment of the included studies was undertaken in this study, as the purpose of a scoping review is to comprehensively gather and map existing evidence from the literature, summarize key concepts, and identify available findings across different contexts and countries that can inform and guide future strategies.

### Data Synthesis

Data extracted from both screening levels were categorized into input, throughput, and output components by KASIP, LKY, MK, and NJN, based on Asplin et al [14]. Input refers to factors that drive demand for ED services, including emergency, unscheduled, and safety-net care. Throughput encompasses ED activities such as triage, diagnostic evaluation, treatment, and inpatient boarding, while output addresses challenges in transferring admitted patients to inpatient beds, including limited follow-up care, bed shortages, or transfer delays. Using this concept, specific indicators of ED overcrowding, such as waiting time, length of stay, bed occupancy rate, and patient flow, were systematically extracted and organized to identify the most frequently used measures globally. This approach helps identify the bottlenecks, allowing the appropriate indicators and the planning of targeted strategies to improve the affected areas.

## Results

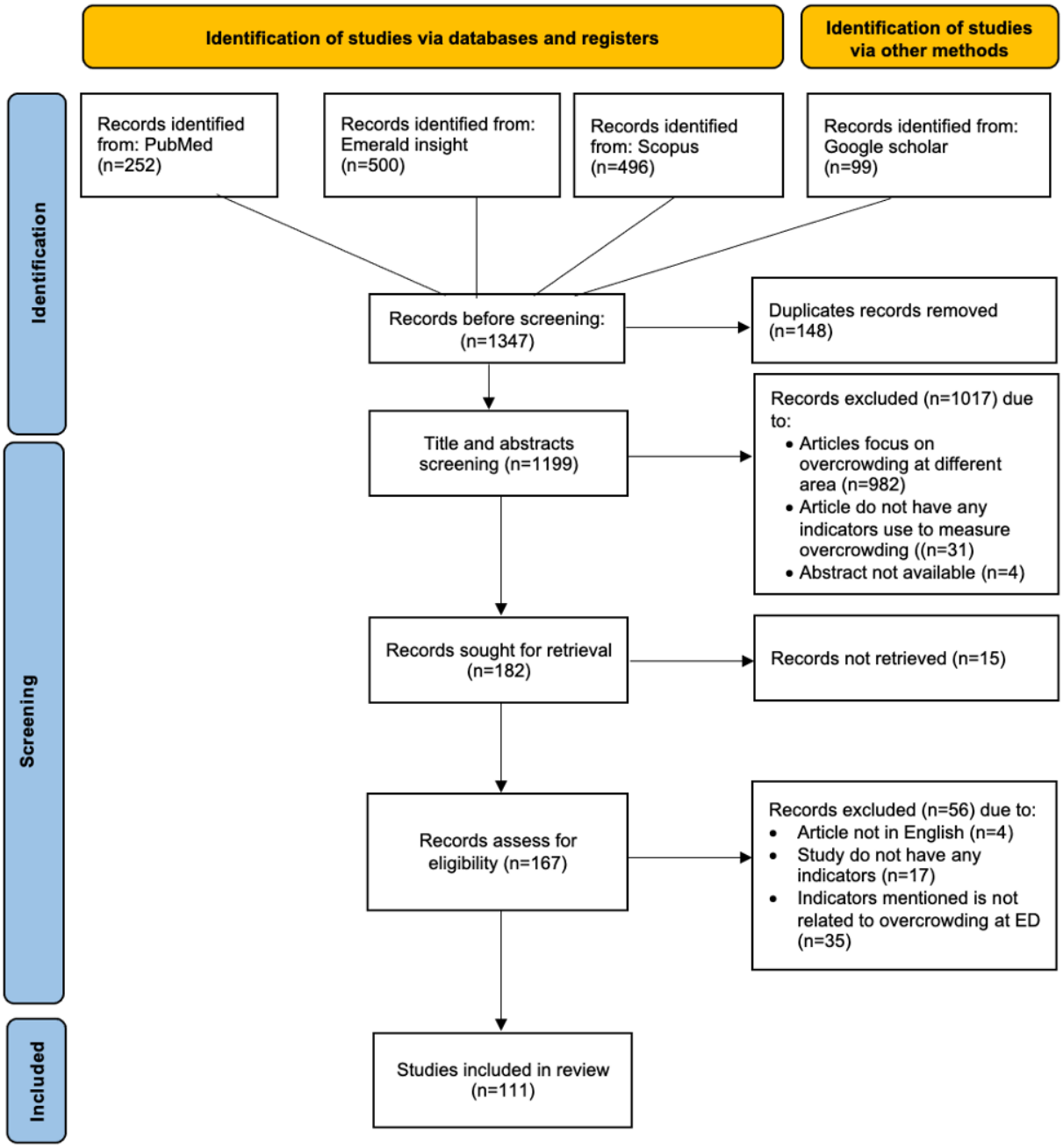
### Study Selection

Figure 1 illustrates the process of identifying, screening, and inclusion of studies for the review. Initially, a total of 1347 records were identified from 4 databases: PubMed (n=252), Emerald Insight (n=500), Scopus (n=496), and Google Scholar (n=99). After removal of 148 duplicate records, 1199 titles and abstracts were screened. Of these, 1017 were excluded for the following reasons: focus on overcrowding in other areas (n=982), absence of indicators measuring overcrowding (n=31), or unavailability of abstract (n=4). Subsequently, 182 records were sought for retrieval, but 15 were unable to be retrieved. Next, 167 records were assessed for eligibility based on inclusion and exclusion criteria, of which 56 were excluded as the article was not in English (n=4), study did not report any indicators (n=17), or the indicators were not related to ED overcrowding (n=35). In the final level, 111 studies [3,7,11,15-121] as mentioned in the List of indicators in Multimedia Appendix 2, were included for data synthesis.



XSL•FO
RenderX

**Figure 1.** Flowchart of the article selection. ED: emergency department.




## Characteristics of the Included Studies

The greatest number of studies examining indicators was published in 2014 (n=14) [15-27, 109] followed by 2019 and 2021 (n=14 each) [1,3,28-51,122], 2017 (n=13) [52-63,123], 2016 and 2018 (n= 9 each) [7,64-80], 2020 (n=11) [81-91], 2022 (n=9) [92-100], and 2013 (n=6) and 2015 (n=6) each) [11,101,103-108,110-113]. Studies mostly originated from the United States of America (n=32) [16-18,22,24,28,34,39,40,44,51-54,68,72,77,80,82,83,87,89,95,103,104,107,109,112,114-117] [27]. Additional articles included in this study came from countries such as Australia (n=12) [23,26,31,36,47,55,56,73,75,76,98,122], Canada (n=7) [20,32,45,57,58,78,85], Italy (n=6) [29,37,43,59,92,99], The Netherlands (n=6) [7,33,79,101,111,117], Taiwan (n=6) [3,60,69,88,94,106,118], Sweden (n=4) [35,38,49,100], South Korea (n=5) [25,81,86,105,113], France (n=4) [11,30,71,95], India (n=2) [1,41], Israel (n=3) [48,95,108], Turkey (n=2) [15,19], Brazil (n=2) [42,52], England (n=1) [70], Germany (n=2) [21,93], Saudi Arabia (n=2) [46,90], Switzerland (n=2) [119,120], Belgium (n=1) [123], Iran (n=1) [61], and Vietnam (n=1) [50]. In some studies, the country was not specifically mentioned or not applicable (n=9) [62-65,67,74,91,102,106].

The study designs of the articles primarily included quality improvement, case-control, case study, cohort study, simulation modeling, cross-sectional study, quantitative study, editorial, literature or systematic or scoping review, mixed methods,



XSL•FO
RenderX

retrospective study, quasi-experimental study, and time series. For the study design, most studies used a cross-sectional study design (n = 55) [1,5,7,9,11,12,15,17,19,21,23,26,28,30,33,35,38,41,44,49,52,55,57,59,61,63,66,67,70,73,77,78,81,83,86,89,91,93,96,98,100,101,103,105,107,110,112,114,118,121] followed by a cohort (n=18) [7,11,20,28,31,32,45,46,48,53,54,56,60,69,79,89,112,116]. Detailed characteristics of the studies included in this review are shown in Multimedia Appendix 2.

## Result of Indicator Retrieval

Overall, 111 studies [3,7,11,15-121] reported indicators for measuring overcrowding, either alone or in combination with other tools, while 6 studies used only specific tools without reporting indicators. The list of indicators mentioned in Multimedia Appendix 2, a total of 360 indicators were initially extracted. The minimum number of indicators reported in the studies was one, whereas the maximum number of indicators identified in a single study was 11. Indicators that conveyed similar information and descriptions were identified and merged. Following detailed discussions among the investigators, indicators that conveyed similar information or measured the same underlying concept were grouped together to minimize redundancy. This results in a reduced set of 307 indicators.

These indicators were subsequently analyzed and categorized according to Asplin et al [14] conceptual model of ED overcrowding. Overall, out of 307 indicators identified, the majority, 209 (68%), were classified within the throughput component, reflecting their central role in managing ED performance. In comparison, 62 (20%) indicators were categorized in the output component, while the least proportion, 36 (12%) indicators, were categorized as input. Among the 307 indicators, several of them convey and measured the same parameters. They were then consolidated once more in the final phase of the analysis and presented as a single description for clarity and easier stakeholder comprehension. This process produced 26 distinct indicators.

## Synthesis of Results

The findings indicate that the widely used indicator to measure overcrowding was the strengthening of throughput processes, with ED length of stay (ED LOS) emerging as the most frequently used indicator. This was followed by waiting time and left without being seen (LWBS), each reported in 30 studies [19,20,22-24,28,30,34,43,54,57,65-67,70,73,85,102,104,110,114,115]. Within the input component, the average number of patients visiting the ED was the most commonly applied indicator. For output, ED boarding time was most frequently used. The results from the search strategy for indicators using PubMed, Emerald Insight, Google Scholar, and Scopus, and the frequency of indicator uses across studies, along with their classification according to ED components, are summarized in Table 2.



XSL•FO
RenderX

**Table .** List of indicators with frequency, and ED[a] components classification.

| Indicators | Description | Frequency mentioned (n) | Study | ED components |
|---|---|---|---|---|
| Average number of patients visiting ED | Average total number of patients who attended ED [91,114] | 23 | [7,15,16,27-29,41-43,57,62,64,73,81,101-103,114] | Input |
| Representations to ED within 48 - 72 hours | Percentage of patients making an unplanned visit to ED within 48 - 72 hours after first visit for the same chief complaint [119] | 7 | [15,65,66,104] | Input |
| Ambulance diversion or turnover time | Period during which ED-requests ambulances that would normally bring patients to the hospital to instead proceed to other hospitals presumed to be less crowded [52,105] | 6 | [17,30,31,55,102,106] | Input |
| ED length of stay | Measured from first clinician contact in the ED (triage or ED reception) until the patient physically left the ED (admitted, transferred, or discharged) [2,11,44,61] | 84 | [3,10,12,22,23,28,34,44,46,48,68,69,72,91,101,104,114,115] | Throughput |
| Waiting time | Total time from initial registration or triage to first being seen by a doctor or clinician [20,111] | 31 | [7,10,28,34,44,45,55,56,66,67,72,75,85,91,101,104,114,120] [100] | Throughput |
| Left without being seen | The number of patients who arrived, registered, and underwent the triage process but leave prior to being seen by a medical provider or physician [2,17,74,92] | 30 | [9,12,24,30,34,45,56,66,67,73,85,101,104,114,115] | Throughput |
| ED occupancy rate | The number of occupied beds divided by the total number of ED beds, expressed in percentage [15,24,88,92,112] | 21 | [25,33,35,41,44,46,73,89,92,99,101,105,112,113] [100] | Throughput |
| Arrival to consultation | Time taken by patient from arrival at primary triage until patient sees a doctor or first consultation by physician [30,65] | 11 | [11,30,58,72,83,85,87,97,103,107,111] | Throughput |
| Time to admit and treat | Time interval spent by patient between arrival in the ED and his or her first consultation by a physician [23,49,80,89,92] | 7 | [16,18,29,39,44,51,55] | Throughput |
| Time to decision and referral | Time measured from patient arrival to decision of referral, discharge, or admission [20,23] | 5 | [11,48,55,60,121] | Throughput |
| Ratio of doctors/patient, nurse/patient | Number of patients each unique clinician is responsible for in a given time slot, presented for each profession [39] | 5 | [38,92] | Throughput |



XSL•FO
RenderX

| Indicators | Description | Frequency mentioned (n) | Study | ED components |
|---|---|---|---|---|
| Mortality rate | Refers to a patient's death when that patient was still physically in ED or had been initially admitted to the hospital from ED, expressed in number [74] | 4 | [35,66,91,104] | Throughput |
| Patient satisfaction | Total percentage of patients ranked their overall ED care as very good or excellent [17,34] | 4 | [17,34] | Throughput |
| Diagnostic turnaround time and time after diagnostic test | Durations of the different diagnostics that a patient is subjected to during their stay in the ED. Defined as the time from order entry to the time the radiology and laboratory releases a report in the system [51] | 3 | [33,115] | Throughput |
| Percentage of patients left against medical advice | Patients who left without completion of medical care [15,111] | 2 | [41,83] | Throughput |
| Time to triage | Measured from patient's arrival to first encounter with triage nurse [11,23] | 2 | [11,48] | Throughput |
| Number of ED beds | Number of licensed ED beds, as presented to state health office [92,94] | 1 | [16] | Throughput |
| Patient throughput | Number of patients who are treated and discharged from a health care facility within a specific period [93] | 1 | [18] | Throughput |
| ED boarding time | Time duration between a patient's bed-request and admission to an inpatient bed [31,33,45,73,97,105] | 25 | [7,10,13,22,29,34,41,44,45,55,56,63,67,73,78,85,86,89,92,114,115] | Output |
| Number of admissions | Number of admissions from ED, represent workload level in ED [121]. | 18 | [16,22,26,29,42-44,51,52,73,88,92,98,114] | Output |
| Discharge time | Time taken from decision to discharge by physician until patient leaves the bed [30] | 9 | [11,26,28,50,51,66,97] | Output |
| Percentage of patients exiting the ED within 4 - 8 hours | Proportion of all patients exiting ED within 4 - 8 hours, either discharged directly from ED or admitted to inpatient wards [5,119] | 7 | [11,18,23,56,62,91,111] | Output |
| Hospital Bed Occupancy Rate (BOR) | Ratio of occupied beds to census beds ratio [85,89]<br>The number of patients in licensed bed and overflow locations over the number of licensed treatment beds [26] | 5 | [18,31,52,62,102] | Output |
| Bed turnover time | Time from when a patient is discharged until an inpatient bed is available for the next patient [3] | 2 | [52] | Output |



XSL•FO
RenderX

| Indicators | Description | Frequency mentioned (n) | Study | ED components |
|---|---|---|---|---|
| Bed waiting time | Time from decision by ED doctor for admission and referral to primary team until patient arrives at bed in ward [30] | 2 | [7,97] | Output |
| Number of hospital beds | Number of licensed hospital beds, as presented to state health office [92,94] | 1 | [16] | Output |

[a]ED: emergency department.

These 26 indicators' descriptions were collated and summarized as seen in Table 2. Categorization into the 3 components of ED crowding showed that the largest proportion of these indicators was concentrated within the throughput component, comprising 15 indicators, followed by the output component with 8 indicators, while the input component accounted for only 3 indicators. These results highlight the predominant representation of throughput elements compared to input and output elements, underscoring their central importance within the framework.

## Discussion

### Principal Results and Comparison With Prior Work

In this review, the numerous indicators used to assess overcrowding in the EDs illustrated the complexities and diversity of the ED system around the world. The methods by which these indicators are categorized also revealed differences in the conceptualization of overcrowding in the literature. Based on the framework by Asplin et al [14], the indicators obtained from the review were mapped to the input, throughput, and output components. Most of the indicators obtained in this review measured the throughput component, highlighting the importance of examining internal ED work processes, including triaging, room placement, initial provider evaluation, diagnostic testing, and treatment [7].

Throughput indicators dominate the ED performance system largely due to regulatory mandates, policy initiatives, and their alignment with multiple stakeholder interests. The Centers for Medicare and Medicaid Services has specifically made their metrics related to throughput measures for compliance and reimbursement purposes [124]. Furthermore, these metrics are relatively easy to capture from administrative data and electronic health records without the need for resource-intensive data collection methods [1,125]. Hospital administrators find the throughput metrics relevant due to their interest in operational efficiency and capacity management. Patients and families experience visible evidence of emergency performance through waiting times and length of stay, which made throughput measures the dominant focus compared to other quality dimensions [2,125].

Throughput components are mainly involved in internal processes in the ED. The overcrowding causes are due to the need of the specialists or consultants, a lack of instrumental diagnostic support, and a high number of procedures conducted, which results in a burden in ED and compromises the patient flow [4]. The strength of the manpower and adding resources could help in overcoming the overcrowding issues and fulfill the demand. The strategies from micro-level and macro-level should be managed appropriately based on the local setting and the health care institution’s capability [3,4,126].

Among the 3 components, ED LOS from the throughput category emerged as the most commonly measured indicator across studies. Introduced as part of the National Health Service ED performance metrics in the United Kingdom, length of stay serves as a critical measure of throughput and overcrowding, reflecting delays in diagnostics and treatment. Research has shown that patients discharged during shifts with a mean Length of Stay ≥6 h face a 79% higher risk of 7-day mortality compared to those with a length of stay <1 hour [117]. Prolonged length of stay underscores systemic issues, such as diagnostic delays and bed shortages, highlighting its impact on patient outcomes. Reducing length of stay is vital for improving patient satisfaction and increasing the ED’s capacity to manage higher patient volumes. Prolonged length of stay is associated with higher patient dissatisfaction with ED care. This finding is supported by a study by Parker and Marco [127], which explored the correlation between length of stay and patient satisfaction. The authors discovered that patients were dissatisfied with ED services when the Length of Stay exceeded 195 minutes, whereas they were satisfied when the length of stay was less than 150 minutes.

The throughput indicators, often associated with regulatory and accreditation requirements, are integral to ensuring compliance and improving health care delivery standards [7]. Among these, waiting and boarding times have been recommended as routine quality measures. Waiting time, in particular, is a critical indicator with significant implications for patient satisfaction and perceived quality of care. A prospective cross-sectional study involving 644 patients found that increased waiting room time was associated with perceptions of compromised care (OR [odds ratio] 1.05, 95% CI 1.02 - 1.09 for each additional 10 minutes spent in the waiting room), as well as receiving care in hallways (OR 2.02, 95% CI 1.12 - 3.68) [41]. These perceptions can affect patient trust and the overall reputation of health care facilities [93]. Furthermore, prolonged waiting times can lead to dissatisfaction and health deterioration, potentially resulting in more severe complications. As a visible aspect of overcrowding, waiting time plays a pivotal role in shaping public perception and trust in health care services.

LWBS and ED occupancy rates are good proxies of patient satisfaction [11,18]. The LWBS rate reflects patient frustration and the ED’s inability to manage patient flow effectively.



XSL•FO
RenderX

Studies have consistently documented a positive correlation between ED overcrowding and LWBS rates [11,32,38,93,96]. The number of patients leaving without receiving care varied widely, ranging from 213 over an 18-day period [18] to 14,170 over 27 months [28]. This trend highlights systemic inefficiencies within EDs that compromise patient safety and satisfaction [16]. High LWBS rates indicate long wait times and overcrowded conditions, posing significant challenges to operational efficiency, patient safety, and care quality.

Given the impact of patient satisfaction, EDs are increasingly focusing on improving internal processes and addressing overcrowding. This shift has driven the measurement of additional throughput indicators to inform strategies for enhancing patient flow and overall service quality [19]. This explains why these indicators are frequently measured for overcrowding in this review. In contrast, only 12% (36/307) of the indicators fall under the input component, with the average number of patients visiting the ED being the most common. Likewise, 20% (62/307) of indicators were categorized as output like ED boarding time and number of admissions. Since input indicators are related to the demand for ED services, and output indicators are influenced by the inefficient disposition of ED patients, these processes are more challenging to control. Some studies have identified patient arrival characteristics as influences of input indicators, while the inability to transfer patients out of ED affects output indicators [65,92]. These factors are outside the control of the ED and involve other departments, thus requiring greater coordination.

While throughput measures are important, they must be balanced against input and output indicators to avoid a narrow assessment on the performance of ED overcrowding. Since throughput measures primarily emphasize speed, they risk incentivizing premature discharges to meet time targets. In systems dominated by throughput metrics, other critical dimensions such as clinical quality and patient safety indicators may be overlooked [5,128,129].

## Limitations

This review has several limitations. First, the literature search was conducted in June 2023; additional studies may have been missed. Despite broad eligibility criteria, some relevant studies (eg, non-English or without full-text) may have been excluded. Nevertheless, the compilation and categorization of indicators remain relevant. Second, a key challenge was the lack of clear definitions for certain indicators, such as distinguishing between waiting time and the interval from arrival to consultation. However, the researchers have clearly defined the variables in the results to assist the reader in better understanding. Third, the heterogeneity of indicators across studies also made direct comparisons difficult. Fourth, as no single "perfect" indicator exists, hospital systems may benefit more from using a combination of measures, such as ED occupancy rate, length of stay, and ED volume, to obtain a comprehensive assessment. Fifth, the variability in indicator use across settings complicates efforts to establish universal benchmarks. Sixth, intergroup reliability was not assessed, as the primary objective in this study was to identify indicators and map the findings according to input, throughput, and output components. Methodological rigor was maintained through independent screening, structured consensus discussions, and resolution by a third reviewer. These limitations highlight the need for more standardized definitions and consensus on evidence-based indicators to support meaningful comparisons and guide public health policies.

## Conclusions

In conclusion, the findings of this review highlight that throughput indicators were more prominently studied as key metrics in measuring ED overcrowding. The review provides further evidence that length of stay emerged as a critical indicator reflecting systemic bottlenecks and operational inefficiencies.

These findings underscore the importance of prioritizing throughput indicators, particularly ED LOS, as a benchmark for monitoring and addressing overcrowding. However, the lack of standardized definitions and variability in indicator use highlight a critical need for future research to develop a universally accepted framework for measuring ED overcrowding. Further studies should also explore the integration of input, throughput, and output indicators in predictive models to provide a more holistic understanding of patient flow dynamics.

From a policy perspective, the evidence suggests that interventions targeting diagnostic delays, staffing adequacy, and bed availability should be prioritized to reduce bottlenecks and improve patient outcomes. Policymakers and health care administrators may consider adopting composite measures, rather than relying on a single indicator, to guide resource allocation and performance evaluation. At the national level, embedding standardized overcrowding metrics into hospital accreditation and reporting systems could enhance comparability across institutions and support evidence-based policy development.

Overall, addressing ED overcrowding requires coordinated efforts at both institutional and policy levels. By strengthening measurement frameworks and aligning them with local contexts, health care systems can not only improve patient satisfaction and safety but also build greater resilience in responding to increasing health care demands.

## Acknowledgments

The authors would like to thank the director of our institute for Health Management (IHM), Noraziani Khamis, for granting permission to conduct the study, providing guidance and support throughout, and assisting with the review of the manuscript. The study was conducted as part of our Institute's project which was registered with the National Medical Research Registry (NMRR), Malaysia, under NMRR ID-23-01709-HDO and not funded.



XSL•FO
RenderX

## Funding

The authors declare no financial support was received for this work.

## Authors' Contributions

Conceptualization: NNMN, KASIP, NFS, NJN, MK, ZL, NNRA, LKY, and ISS
Data curation: KASIP
Formal analysis: KASIP, ZL, NJN, and NNMN
Investigation: KASIP, MK, LKY, and NNRA
Methodology: NNMN and MK
Project administration: NNMN
Resources: KASIP and NJN
Supervision: ISS
Validation: KASIP, ISS, LKY, and NJN
Visualization: KASIP, NNMN, and ISS
Writing: NNMN
Writing: NNMN, KASIP, NFS, NJN, MK, ZL, NNRA, LKY, and ISS

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary table of the result from search strategy for indicators.
[PNG File, 205 KB - i-jmr_v15i1e78073_app1.png ]

Multimedia Appendix 2
List of indicators.
[PDF File, 608 KB - i-jmr_v15i1e78073_app2.pdf ]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 312 KB - i-jmr_v15i1e78073_app3.pdf ]

## References

1. Sharma R, Prakash A, Chauhan R, Dhibar DP. Overcrowding an encumbrance for an emergency health-care system: a perspective of health-care providers from tertiary care center in Northern India. J Educ Health Promot 2021;10(1):5. [doi: 10.4103/jehp.jehp_289_20] [Medline: 33688514]
2. Savioli G, Ceresa IF, Gri N, et al. Emergency department overcrowding: understanding the factors to find corresponding solutions. J Pers Med 2022 Feb 14;12(2):279. [doi: 10.3390/jpm12020279] [Medline: 35207769]
3. Hsu CM, Liang LL, Chang YT, Juang WC. Emergency department overcrowding: quality improvement in a Taiwan Medical Center. J Formos Med Assoc 2019 Jan;118(1 Pt 1):186-193. [doi: 10.1016/j.jfma.2018.03.008] [Medline: 29665984]
4. Sartini M, Carbone A, Demartini A, et al. Overcrowding in emergency department: causes, consequences, and solutions-a narrative review. Healthcare (Basel) 2022 Aug 25;10(9):1625. [doi: 10.3390/healthcare10091625] [Medline: 36141237]
5. Bernstein SL, Aronsky D, Duseja R, et al. The effect of emergency department crowding on clinically oriented outcomes. Acad Emerg Med 2009 Jan;16(1):1-10. [doi: 10.1111/j.1553-2712.2008.00295.x] [Medline: 19007346]
6. Jones P, Shepherd M, Wells S, Le Fevre J, Ameratunga S. Review article: what makes a good healthcare quality indicator? A systematic review and validation study. Emerg Med Australas 2014 Apr;26(2):113-124. [doi: 10.1111/1742-6723.12195] [Medline: 24707999]
7. Van Der Linden MC, Van Loon M, Gaakeer MI, Richards JR, Derlet RW, Van Der Linden N. A different crowd, a different crowding level? The predefined thresholds of crowding scales may not be optimal for all emergency departments. Int Emerg Nurs 2018 Nov;41:25-30. [doi: 10.1016/j.ienj.2018.05.004] [Medline: 29880260]
8. Jamiso DT, et al. Disease Control Priorities in Developing Countries: World Bank Publications; 2006. [Medline: 21250309]
9. Crawford K, Morphet J, Jones T, Innes K, Griffiths D, Williams A. Initiatives to reduce overcrowding and access block in Australian emergency departments: a literature review. Collegian 2014;21(4):359-366. [doi: 10.1016/j.colegn.2013.09.005] [Medline: 25632734]
10. Ospina MB, Bond K, Schull M, Innes G, Blitz S, Rowe BH. Key indicators of overcrowding in Canadian emergency departments: a Delphi study. CJEM 2007 Sep;9(5):339-346. [doi: 10.1017/s1481803500015281] [Medline: 17935649]



XSL•FO
RenderX

11. Casalino E, Choquet C, Bernard J, et al. Predictive variables of an emergency department quality and performance indicator: a 1-year prospective, observational, cohort study evaluating hospital and emergency census variables and emergency department time interval measurements. Emerg Med J 2013 Aug;30(8):638-645. [doi: 10.1136/emermed-2012-201404] [Medline: 22906702]
12. Nyanchoka L, Tudur-Smith C, Thu VN, Iversen V, Tricco AC, Porcher R. A scoping review describes methods used to identify, prioritize and display gaps in health research. J Clin Epidemiol 2019 May;109:99-110. [doi: 10.1016/j.jclinepi.2019.01.005] [Medline: 30708176]
13. Furuya-Kanamori L, Lin L, Kostoulas P, Clark J, Xu C. Limits in the search date for rapid reviews of diagnostic test accuracy studies. Res Synth Methods 2023 Mar;14(2):173-179. [doi: 10.1002/jrsm.1598] [Medline: 36054082]
14. Asplin BR, Magid DJ, Rhodes KV, Solberg LI, Lurie N, Camargo CA. A conceptual model of emergency department crowding. Ann Emerg Med 2003 Aug;42(2):173-180. [doi: 10.1067/mem.2003.302] [Medline: 12883504]
15. Erenler AK, Akbulut S, Guzel M, et al. Reasons for overcrowding in the emergency department: experiences and suggestions of an education and research hospital. Turk J Emerg Med 2014 Jun;14(2):59-63. [doi: 10.5505/1304.7361.2014.48802] [Medline: 27331171]
16. Weiss SJ, Rogers DB, Maas F, Ernst AA, Nick TG. Evaluating community ED crowding: the Community ED overcrowding scale study. Am J Emerg Med 2014 Nov;32(11):1357-1363. [doi: 10.1016/j.ajem.2014.08.035] [Medline: 25234796]
17. Patel PB, Combs MA, Vinson DR. Reduction of admit wait times: the effect of a leadership-based program. Acad Emerg Med 2014 Mar;21(3):266-273. [doi: 10.1111/acem.12327] [Medline: 24628751]
18. Di Somma S, Paladino L, Vaughan L, Lalle I, Magrini L, Magnanti M. Overcrowding in emergency department: an international issue. Intern Emerg Med 2015 Mar;10(2):171-175. [doi: 10.1007/s11739-014-1154-8] [Medline: 25446540]
19. Aksel G, Bildik F, Demircan A, et al. Effects of fast-track in a university emergency department through the National Emergency Department Overcrowding Study. J Pak Med Assoc 2014 Jul;64(7):791-797. [Medline: 25255588]
20. Doan Q, Genuis ED, Yu A. Trends in use in a Canadian pediatric emergency department. CJEM 2014 Sep;16(5):405-410. [doi: 10.2310/8000.2013.131280] [Medline: 25227649]
21. Korsten P, Sliwa B, Kühn M, Müller GA, Blaschke S. Impact of professional quality management on interdisciplinary emergency care units. Eur J Emerg Med 2014 Apr;21(2):98-104. [doi: 10.1097/MEJ.0b013e3283613e8d] [Medline: 24573191]
22. Handel DA, Fu R, Vu E, et al. Association of emergency department and hospital characteristics with elopements and length of stay. J Emerg Med 2014 Jun;46(6):839-846. [doi: 10.1016/j.jemermed.2013.08.133] [Medline: 24462026]
23. Sullivan CM, Staib A, Flores J, et al. Aiming to be NEAT: safely improving and sustaining access to emergency care in a tertiary referral hospital. Aust Health Rev 2014 Nov;38(5):564-574. [doi: 10.1071/AH14083] [Medline: 25297518]
24. Wang H, Robinson RD, Bunch K, et al. The inaccuracy of determining overcrowding status by using the national ED overcrowding study tool. Am J Emerg Med 2014 Oct;32(10):1230-1236. [doi: 10.1016/j.ajem.2014.07.032] [Medline: 25176566]
25. Jo S, Jin YH, Lee JB, Jeong T, Yoon J, Park B. Emergency department occupancy ratio is associated with increased early mortality. J Emerg Med 2014 Feb;46(2):241-249. [doi: 10.1016/j.jemermed.2013.05.026] [Medline: 23992849]
26. Khanna S, Boyle J, Zeitz K. Using capacity alert calls to reduce overcrowding in a major public hospital. Aust Health Rev 2014 Jun;38(3):318-324. [doi: 10.1071/AH13103] [Medline: 24814040]
27. Handel DA, French LK, Nichol J, Momberger J, Fu R. Associations between patient and emergency department operational characteristics and patient satisfaction scores in an adult population. Ann Emerg Med 2014 Dec;64(6):604-608. [doi: 10.1016/j.annemergmed.2014.07.451] [Medline: 25182541]
28. Xu Y, Ho V. Freestanding emergency departments in Texas do not alleviate congestion in hospital-based emergency departments. Am J Emerg Med 2020 Mar;38(3):471-476. [doi: 10.1016/j.ajem.2019.05.020] [Medline: 31126669]
29. Strada A, Bravi F, Valpiani G, Bentivegna R, Carradori T. Do health care professionals' perceptions help to measure the degree of overcrowding in the emergency department? A pilot study in an Italian University hospital. BMC Emerg Med 2019 Aug 27;19(1):47. [doi: 10.1186/s12873-019-0259-9] [Medline: 31455226]
30. Ajmi F, Zgaya H, Othman SB, Hammadi S. Agent-based dynamic optimization for managing the workflow of the patient's pathway. Simul Model Pract Theory 2019 Nov;96:101935. [doi: 10.1016/j.simpat.2019.101935]
31. Forero R, Man N, Ngo H, et al. Impact of the four‐hour National Emergency Access Target on 30 day mortality, access block and chronic emergency department overcrowding in Australian emergency departments. Emerg Medicine Australasia 2019 Feb;31(1):58-66. [doi: 10.1111/1742-6723.13151] [Medline: 30062847]
32. Doan Q, Wong H, Meckler G, et al. The impact of pediatric emergency department crowding on patient and health care system outcomes: a multicentre cohort study. CMAJ 2019 Jun 10;191(23):E627-E635. [doi: 10.1503/cmaj.181426] [Medline: 31182457]
33. van der Linden MCC, van Ufford H, Project Group Medical Specialists, van der Linden NN. The impact of a multimodal intervention on emergency department crowding and patient flow. Int J Emerg Med 2019 Aug 27;12(1):21. [doi: 10.1186/s12245-019-0238-7] [Medline: 31455260]
34. Spencer S, Stephens K, Swanson-Biearman B, Whiteman K. Health care provider in triage to improve outcomes. J Emerg Nurs 2019 Sep;45(5):561-566. [doi: 10.1016/j.jen.2019.01.008] [Medline: 30827577]




XSL•FO
RenderX

35. Berg LM, Ehrenberg A, Florin J, Östergren J, Discacciati A, Göransson KE. Associations between crowding and ten-day mortality among patients allocated lower triage acuity levels without need of acute hospital care on departure from the emergency department. Ann Emerg Med 2019 Sep;74(3):345-356. [doi: 10.1016/j.annemergmed.2019.04.012] [Medline: 31229391]
36. Boyle LM, Marshall AH, Mackay M. Coxian phase-type regression models for understanding the relationship between patient attributes, overcrowding, and length of stay in hospital emergency departments. : Springer; 2020 Presented at: 4th International Conference on Health Care Systems Engineering, HCSE 2019; May 30 to Jun 1, 2019; Montréal, Canada p. 53-64. [doi: 10.1007/978-3-030-39694-7_5]
37. Torto AD, Pozzi R, Porazzi E, Garagiola E, Strozzi F. Length of stay reduction in the emergency department and its quantification using complex network theory. IJOR 2019;36(3):337. [doi: 10.1504/IJOR.2019.103122]
38. Berg LM, Ehrenberg A, Florin J, Östergren J, Göransson KE. Significant changes in emergency department length of stay and case mix over eight years at a large Swedish University Hospital. Int Emerg Nurs 2019 Mar;43:50-55. [doi: 10.1016/j.ienj.2018.08.001] [Medline: 30190224]
39. Acuna JA, Zayas-Castro JL, Charkhgard H. Ambulance allocation optimization model for the overcrowding problem in US emergency departments: a case study in Florida. Socioecon Plann Sci 2020 Sep;71:100747. [doi: 10.1016/j.seps.2019.100747]
40. Davis Z, Zobel CW, Khansa L, Glick RE. Emergency department resilience to disaster‐level overcrowding: a component resilience framework for analysis and predictive modeling. J of Ops Management 2020 Jan;66(1-2):54-66. [doi: 10.1002/joom.1017]
41. Menon NVB, Jayashree M, Nallasamy K, Angurana SK, Bansal A. Bed utilization and overcrowding in a high-volume tertiary level pediatric emergency department. Indian Pediatr 2021 Aug 15;58(8):723-725. [doi: 10.1007/s13312-021-2278-3] [Medline: 33634795]
42. Cirillo W, Freitas LRC, Kitaka EL, et al. Impact of emergency short‐stay unit opening on in‐hospital global and cardiology indicators. Evaluation Clinical Practice 2021 Dec;27(6):1262-1270. [doi: 10.1111/jep.13534] [Medline: 33421284]
43. Savioli G, Ceresa IF, Novelli V, Ricevuti G, Bressan MA, Oddone E. How the coronavirus disease 2019 pandemic changed the patterns of healthcare utilization by geriatric patients and the crowding: a call to action for effective solutions to the access block. Intern Emerg Med 2022 Mar;17(2):503-514. [doi: 10.1007/s11739-021-02732-w] [Medline: 34106397]
44. Gorski JK, Arnold TS, Usiak H, Showalter CD. Crowding is the strongest predictor of left without being seen risk in a pediatric emergency department. Am J Emerg Med 2021 Oct;48:73-78. [doi: 10.1016/j.ajem.2021.04.005] [Medline: 33845424]
45. Kegel F, Luo OD, Richer S. The impact of extreme heat events on emergency departments in Canadian hospitals. Wilderness Environ Med 2021 Dec;32(4):433-440. [doi: 10.1016/j.wem.2021.05.004] [Medline: 34364750]
46. Al-Qahtani MF, Khubrani FY. Exploring potential association between emergency department crowding status and patients' length of stay at a university hospital in Saudi Arabia. Open Access Emerg Med 2021;13:257-263. [doi: 10.2147/OAEM.S305885] [Medline: 34188561]
47. Kenny E, Hassanzadeh H, Khanna S, Boyle J, Louise S. Patient flow simulation using historically informed synthetic data. In: Studies in Health Technology and Informatics: IOS Press; 2021, Vol. 276:32-37. [doi: 10.3233/SHTI210007]
48. Trotzky D, Posner L, Mosery J, Cohen A, Avisar S, Pachys G. Do automatic push notifications improve patient flow in the emergency department? Analysis of an ED in a large medical center in Israel. PLoS ONE 2021;16(10):e0258169. [doi: 10.1371/journal.pone.0258169] [Medline: 34618849]
49. Wretborn J, Starkenberg H, Ruge T, Wilhelms DB, Ekelund U. Validation of the modified Skåne emergency department assessment of patient load (mSEAL) model for emergency department crowding and comparison with international models; an observational study. BMC Emerg Med 2021 Feb 22;21(1):21. [doi: 10.1186/s12873-021-00414-6] [Medline: 33618658]
50. Le DX, Do HT, Bui KT, et al. Lean management for improving hospital waiting times-case study of a Vietnamese public/general hospital emergency department. Int J Health Plann Manage 2022 Jan;37(1):156-170. [doi: 10.1002/hpm.3310] [Medline: 34490656]
51. Theodoro D, Vyas N, Ablordeppey E, et al. Central venous catheter adverse events are not associated with crowding indicators. W J Emerg Med 2020;22(2). [doi: 10.5811/westjem.2020.10.48279] [Medline: 33856335]
52. Soares VS. Analysis of the internal bed regulation committees from hospitals of a Southern Brazilian city. Einstein (Sao Paulo) 2017;15(3):339-343. [doi: 10.1590/S1679-45082017GS3878] [Medline: 29091157]
53. Phillips JL, Jackson BE, Fagan EL, et al. Overcrowding and Its association with patient outcomes in a median-low volume emergency department. J Clin Med Res 2017 Nov;9(11):911-916. [doi: 10.14740/jocmr3165w] [Medline: 29038668]
54. Wang H, Ojha RP, Robinson RD, et al. Optimal measurement interval for emergency department crowding estimation tools. Ann Emerg Med 2017 Nov;70(5):632-639. [doi: 10.1016/j.annemergmed.2017.04.012] [Medline: 28688771]
55. Burke JA, Greenslade J, Chabrowska J, et al. Two hour evaluation and referral model for shorter turnaround times in the emergency department. Emerg Med Australas 2017 Jun;29(3):315-323. [doi: 10.1111/1742-6723.12781] [Medline: 28455884]
56. Street M, Mohebbi M, Berry D, Cross A, Considine J. Influences on emergency department length of stay for older people. Eur J Emerg Med 2018 Aug;25(4):242-249. [doi: 10.1097/MEJ.0000000000000452] [Medline: 28151752]




XSL•FO
RenderX

57. Truong M, Meckler G, Doan QH. Emergency department return visits within a large geographic area. J Emerg Med 2017 Jun;52(6):801-808. [doi: 10.1016/j.jemermed.2017.01.009] [Medline: 28228344]
58. Georgio G, Guttmann A, Doan QH. Emergency department flow measures for adult and pediatric patients in British Columbia and Ontario: a retrospective, repeated cross-sectional study. J Emerg Med 2017 Sep;53(3):418-426. [doi: 10.1016/j.jemermed.2017.05.004] [Medline: 28676415]
59. Landa P, Sonnessa M, Resta M, Tànfani E, Testi A. A hybrid simulation approach to analyse patient boarding in emergency departments. In: Health Care Systems Engineering: Springer; 2017:133-144. [doi: 10.1007/978-3-319-66146-9_12]
60. Chiu IM, Lin YR, Syue YJ, Kung CT, Wu KH, Li CJ. The influence of crowding on clinical practice in the emergency department. Am J Emerg Med 2018 Jan;36(1):56-60. [doi: 10.1016/j.ajem.2017.07.011] [Medline: 28705743]
61. Rahmani F, Rezazadeh F, Ala A, Soleimanpour M, Mehdizadeh Esfanjani R, Soleimanpour H. Evaluation of overcrowding of emergency department in Imam Reza Hospital in 2015 by implementing 2 scales: NEDOCS and EDWIN. Iran Red Crescent Med J 2017;19(6):1-4. [doi: 10.5812/ircmj.15609]
62. Scott I, Sullivan C, Staib A, Bell A. Deconstructing the 4-h rule for access to emergency care and putting patients first. Aust Health Rev 2018 Dec;42(6):698-702. [doi: 10.1071/AH17083] [Medline: 29032791]
63. Dai JG, Shi P. A two-time-scale approach to time-varying queues in hospital inpatient flow management. Oper Res 2017 Apr;65(2):514-536. [doi: 10.1287/opre.2016.1566]
64. Santos E, Cardoso D, Queirós P, Cunha M, Rodrigues M, Apóstolo J. The effects of emergency department overcrowding on admitted patient outcomes: a systematic review protocol. JBI Database System Rev Implement Rep 2016 May;14(5):96-102. [doi: 10.11124/JBISRIR-2016-002562] [Medline: 27532467]
65. Ming T, Lai A, Lau PM. Can team triage improve patient flow in the emergency department? A systematic review and meta-analysis. Adv Emerg Nurs J 2016;38(3):233-250. [doi: 10.1097/TME.0000000000000113] [Medline: 27482995]
66. Burström L, Engström ML, Castrén M, Wiklund T, Enlund M. Improved quality and efficiency after the introduction of physician-led team triage in an emergency department. Ups J Med Sci 2016;121(1):38-44. [doi: 10.3109/03009734.2015.1100223] [Medline: 26553523]
67. Bucci S, de Belvis AG, Marventano S, et al. Emergency department crowding and hospital bed shortage: is lean a smart answer? A systematic review. Eur Rev Med Pharmacol Sci 2016 Oct;20(20):4209-4219. [Medline: 27831655]
68. Fuentes E, Shields JF, Chirumamilla N, et al. “One-way-street” streamlined admission of critically ill trauma patients reduces emergency department length of stay. Intern Emerg Med 2017 Oct;12(7):1019-1024. [doi: 10.1007/s11739-016-1511-x] [Medline: 27473424]
69. Li CJ, Syue YJ, Lin YR, et al. Influence of CT utilisation on patient flow in the emergency department: a retrospective 1-year cohort study. BMJ Open 2016 May 4;6(5):e010815. [doi: 10.1136/bmjopen-2015-010815] [Medline: 27147387]
70. Boyle A, Abel G, Raut P, et al. Comparison of the International Crowding Measure in Emergency Departments (ICMED) and the National Emergency Department Overcrowding Score (NEDOCS) to measure emergency department crowding: pilot study. Emerg Med J 2016 May;33(5):307-312. [doi: 10.1136/emermed-2014-203616] [Medline: 26739294]
71. Ghanes K, Jouini O, Wargon M, Jemai Z. Modeling and analysis of triage nurse ordering in emergency departments. Presented at: 2015 International Conference on Industrial Engineering and Systems Management (IESM); Oct 21, 2015; Seville, Spain p. 228-235. [doi: 10.1109/IESM.2015.7380163]
72. Cocke S, Guinn D, MacBlane E, et al. UVA emergency department patient flow simulation and analysis. : IEEE; 2016 Presented at: 2016 Systems and Information Engineering Design Symposium (SIEDS); Apr 29, 2016; Charlottesville, VA, USA p. 118-123. [doi: 10.1109/SIEDS.2016.7489282]
73. Ngo H, Forero R, Mountain D, et al. Impact of the four-hour rule in Western Australian hospitals: trend analysis of a large record linkage study 2002-2013. PLoS ONE 2018;13(3):e0193902. [doi: 10.1371/journal.pone.0193902] [Medline: 29538401]
74. Li M, Vanberkel P, Carter AJE. A review on ambulance offload delay literature. Health Care Manag Sci 2019 Dec;22(4):658-675. [doi: 10.1007/s10729-018-9450-x] [Medline: 29982911]
75. Thapa RR, Bhuiyan M, Krishna A, Prasad PWC. Application of RFID technology to reduce overcrowding in hospital emergency departments. In: Advances in Information Systems Development: Methods, Tools and Management: Springer; 2018:17-32. [doi: 10.1007/978-3-319-74817-7_2]
76. Andrews R, Suriadi S, Wynn M, Rothwell S. Improving patient flows at St. Andrew’s War Memorial Hospital’s emergency department through process mining. In: Business Process Management Cases: Digital Innovation and Business Transformation in Practice: Springer; 2018, Vol. Part F612:311-333. [doi: 10.1007/978-3-319-58307-5_17]
77. Wallingford G, Joshi N, Callagy P, Stone J, Brown I, Shen S. Introduction of a horizontal and vertical split flow model of emergency department patients as a response to overcrowding. J Emerg Nurs 2018 Jul;44(4):345-352. [doi: 10.1016/j.jen.2017.10.017] [Medline: 29169818]
78. Salehi L, Phalpher P, Valani R, et al. Emergency department boarding: a descriptive analysis and measurement of impact on outcomes. CJEM 2018 Nov;20(6):929-937. [doi: 10.1017/cem.2018.18] [Medline: 29619913]
79. van der Veen D, Remeijer C, Fogteloo AJ, Heringhaus C, de Groot B. Independent determinants of prolonged emergency department length of stay in a tertiary care centre: a prospective cohort study. Scand J Trauma Resusc Emerg Med 2018 Sep 20;26(1):81. [doi: 10.1186/s13049-018-0547-5] [Medline: 30236125]




XSL•FO
RenderX

80. Molla M, Warren DS, Stewart SL, Stocking J, Johl H, Sinigayan V. A lean six sigma quality improvement project improves timeliness of discharge from the hospital. Jt Comm J Qual Patient Saf 2018 Jul;44(7):401-412. [doi: 10.1016/j.jcjq.2018.02.006] [Medline: 30008352]
81. Kim DU, Park YS, Park JM, et al. Influence of overcrowding in the emergency department on return visit within 72 hours. J Clin Med 2020 May 9;9(5):1406. [doi: 10.3390/jcm9051406] [Medline: 32397560]
82. d'Etienne JP, Zhou Y, Kan C, et al. Two-step predictive model for early detection of emergency department patients with prolonged stay and its management implications. Am J Emerg Med 2021 Feb;40:148-158. [doi: 10.1016/j.ajem.2020.01.050] [Medline: 32063427]
83. Shah R, Leno R, Sinert R. Impact of provider-in-triage in a safety-net hospital. J Emerg Med 2020 Sep;59(3):459-465. [doi: 10.1016/j.jemermed.2020.04.059] [Medline: 32595053]
84. Wachtel G, Elalouf A. Addressing overcrowding in an emergency department: an approach for identifying and treating influential factors and a real-life application. Isr J Health Policy Res 2020 Sep 2;9(1):37. [doi: 10.1186/s13584-020-00390-5] [Medline: 32873328]
85. Peng Q, Yang J, Strome T, Weldon E, Chochinov A. Evaluation of physician in triage impact on overcrowding in emergency department using discrete-event simulation. 105267/jjpm 2020;5(4):211-226. [doi: 10.5267/j.jpm.2020.8.002]
86. Choi Y, Jeong J, Kim BG. Admission Decisions Made by Emergency Physicians Can Reduce the Emergency Department Length of Stay for Medical Patients. Emerg Med Int 2020;2020(1):8392832. [doi: 10.1155/2020/8392832] [Medline: 32104606]
87. Payne AS, Brown KM, Berkowitz D, et al. Improving throughput for mid-acuity patients in the pediatric emergency department. Pediatric Quality & Safety 2020;5(3):e302. [doi: 10.1097/pq9.0000000000000302]
88. Tsai JCH, Weng SJ, Liu SC, et al. Adjusting daily inpatient bed allocation to smooth emergency department occupancy variation. Healthcare (Basel) 2020 Mar 28;8(2):78. [doi: 10.3390/healthcare8020078] [Medline: 32231146]
89. Hoot NR, Banuelos RC, Chathampally Y, Robinson DJ, Voronin BW, Chambers KA. Does crowding influence emergency department treatment time and disposition? J Am Coll Emerg Physicians Open 2021 Feb;2(1):e12324. [doi: 10.1002/emp2.12324] [Medline: 33521777]
90. Khubrani FY, Al-Qahtani MF. Association between emergency department overcrowding and mortality at a teaching hospital in Saudi Arabia. TOPHJ 2020 Dec 31;13(1):756-762. [doi: 10.2174/1874944502013010756]
91. Bélanger V, Lahrichi N, Lanzarone E, Yalçındağ S. Health Care Systems Engineering: Springer; 2020. [doi: 10.1007/978-3-030-39694-7]
92. Marsilio M, Roldan ET, Salmasi L, Villa S. Operations management solutions to improve ED patient flows: evidence from the Italian NHS. BMC Health Serv Res 2022 Jul 30;22(1):974. [doi: 10.1186/s12913-022-08339-x] [Medline: 35908053]
93. Otto R, Blaschke S, Schirrmeister W, Drynda S, Walcher F, Greiner F. Length of stay as quality indicator in emergency departments: analysis of determinants in the German Emergency Department Data Registry (AKTIN registry). Intern Emerg Med 2022 Jun;17(4):1199-1209. [doi: 10.1007/s11739-021-02919-1] [Medline: 34989969]
94. Tseng WM, Lin PH, Wu PC, Kao CH. Examining patient flow in a tertiary hospital's emergency department at a low coronavirus prevalence region. BMC Emerg Med 2022 Jul 27;22(1):137. [doi: 10.1186/s12873-022-00694-6] [Medline: 35896981]
95. Kadri F, Dairi A, Harrou F, Sun Y. Towards accurate prediction of patient length of stay at emergency department: a GAN-driven deep learning framework. J Ambient Intell Humaniz Comput 2022 Feb 3;14(9):1-15. [doi: 10.1007/s12652-022-03717-z] [Medline: 35132336]
96. Badr S, Nyce A, Awan T, Cortes D, Mowdawalla C, Rachoin JS. Measures of emergency department crowding, a systematic review. How to make sense of a long list. Open Access Emerg Med 2022;14:5-14. [doi: 10.2147/OAEM.S338079] [Medline: 35018125]
97. Noris NJ, Putera K, Libasin Z, Krishnan M. Lean healthcare implementation in Malaysian specialist hospitals: challenges and performance evaluation. J Heal Transl Med 2022;25(1):27-39. [doi: 10.22452/jummec.vol25no1.6]
98. Mallows JL. Effects of staff grade, overcrowding and presentations on emergency department performance: a regression model. Emerg Medicine Australasia 2022 Jun;34(3):341-346. [doi: 10.1111/1742-6723.13889]
99. Improta G, Majolo M, Raiola E, Russo G, Longo G, Triassi M. A case study to investigate the impact of overcrowding indices in emergency departments. BMC Emerg Med 2022 Aug 9;22(1):143. [doi: 10.1186/s12873-022-00703-8] [Medline: 35945503]
100. Wretborn J, Ekelund U, B Wilhelms D. Differentiating properties of occupancy rate and workload to estimate crowding: A Swedish national cross-sectional study. J Am Coll Emerg Physicians Open 2022 Feb;3(1):e12648. [doi: 10.1002/emp2.12648] [Medline: 35079734]
101. van der Linden C, Reijnen R, Derlet RW, et al. Emergency department crowding in The Netherlands: managers' experiences. Int J Emerg Med 2013 Oct 24;6(1):1-8. [doi: 10.1186/1865-1380-6-41] [Medline: 24156298]
102. Johnson KD, Winkelman C. The effect of emergency department crowding on patient outcomes: a literature review. Adv Emerg Nurs J 2011;33(1):39-54. [doi: 10.1097/TME.0b013e318207e86a] [Medline: 21317697]
103. de Araujo P, Khraiche M, Tukan A. Does overcrowding and health insurance type impact patient outcomes in emergency departments? Health Econ Rev 2013 Nov 12;3(1):1-7. [doi: 10.1186/2191-1991-3-25] [Medline: 24229451]




XSL•FO
RenderX

104. Wang H, Robinson RD, Garrett JS, et al. Use of the SONET score to evaluate high volume emergency department overcrowding: a prospective derivation and validation study. Emerg Med Int 2015;2015:401757. [doi: 10.1155/2015/401757] [Medline: 26167302]
105. Lee YJ, Shin SD, Lee EJ, Cho JS, Cha WC. Emergency department overcrowding and ambulance turnaround time. PLoS ONE 2015;10(6):e0130758. [doi: 10.1371/journal.pone.0130758] [Medline: 26115183]
106. Lin CH, Kao CY, Huang CY. Managing emergency department overcrowding via ambulance diversion: a discrete event simulation model. J Formos Med Assoc 2015 Jan;114(1):64-71. [doi: 10.1016/j.jfma.2012.09.007] [Medline: 25618586]
107. Konrad R, DeSotto K, Grocela A, et al. Modeling the impact of changing patient flow processes in an emergency department: insights from a computer simulation study. Oper Res Health Care 2013 Dec;2(4):66-74. [doi: 10.1016/j.orhc.2013.04.001]
108. Elalouf A, Wachtel G. An alternative scheduling approach for improving patient-flow in emergency departments. Oper Res Health Care 2015 Dec;7:94-102. [doi: 10.1016/j.orhc.2015.08.002]
109. Kim BBJ, Delbridge TR, Kendrick DB. Improving process quality for pediatric emergency department. Int J Health Care Qual Assur 2014;27(4):336-346. [doi: 10.1108/IJHCQA-11-2012-0117] [Medline: 25076607]
110. Fayyaz J, Khursheed M, Mir MU, Mehmood A. Missing the boat: odds for the patients who leave ED without being seen. BMC Emerg Med 2013 Jan 16;13(1):1-9. [doi: 10.1186/1471-227X-13-1] [Medline: 23324162]
111. Vegting IL, Alam N, Ghanes K, et al. What are we waiting for? Factors influencing completion times in an academic and peripheral emergency department. Neth J Med 2015 Aug;73(7):331-340. [Medline: 26314716]
112. Tekwani KL, Kerem Y, Mistry CD, Sayger BM, Kulstad EB. Emergency department crowding is associated with reduced satisfaction scores in patients discharged from the emergency department. West J Emerg Med 2013 Feb;14(1):11-15. [doi: 10.5811/westjem.2011.11.11456] [Medline: 23447751]
113. Kang J, Kim J, Jo YH, et al. ED crowding and the outcomes of out-of-hospital cardiac arrest. Am J Emerg Med 2015 Nov;33(11):1659-1664. [doi: 10.1016/j.ajem.2015.08.002] [Medline: 26324002]
114. Savioli G, Ceresa IF, Bressan MA, et al. Five level triage vs. four level triage in a quaternary emergency department: national analysis on waiting time, validity, and crowding-the CREONTE (Crowding and RE-Organization National TriagE) study group. Medicina (Kaunas) 2023 Apr 17;59(4):781. [doi: 10.3390/medicina59040781] [Medline: 37109739]
115. Watson A, Stuart WP. Improving safety and quality with an emergency department overcrowding plan. J Emerg Nurs 2023 Sep;49(5):680-693. [doi: 10.1016/j.jen.2023.06.002] [Medline: 37409998]
116. Hsieh A, Arena A, Oraha A, et al. Implementation of vertical split flow model for patient throughput at a community hospital emergency department. J Emerg Med 2023 Jan;64(1):77-82. [doi: 10.1016/j.jemermed.2022.10.007] [Medline: 36641257]
117. Van Der Linden MC, Van Loon-Van Gaalen M, Richards JR, Van Woerden G, Van Der Linden N. Effects of process changes on emergency department crowding in a changing world: an interrupted time-series analysis. Int J Emerg Med 2023 Feb 15;16(1):6. [doi: 10.1186/s12245-023-00479-z] [Medline: 36792991]
118. Hsu NC, Shu CC, Lin YF, Yang MC, Su S, Ko WJ. Why do general medical patients have a lengthy wait in the emergency department before admission? J Formos Med Assoc 2014 Aug;113(8):557-561. [doi: 10.1016/j.jfma.2012.08.005] [Medline: 25037761]
119. Hofer KD, Saurenmann RK. Parameters affecting length of stay in a pediatric emergency department: a retrospective observational study. Eur J Pediatr 2017 May;176(5):591-598. [doi: 10.1007/s00431-017-2879-y] [Medline: 28275860]
120. Schmutz T, Le Terrier C, Ribordy V, Guechi Y. No waiting lying in a corridor: a quality improvement initiative in an emergency department. BMJ Open Qual 2023 Aug;12(3):e002431. [doi: 10.1136/bmjoq-2023-002431] [Medline: 37640478]
121. George G, Jell C, Todd BS. Effect of population ageing on emergency department speed and efficiency: a historical perspective from a district general hospital in the UK. Emerg Med J 2006 May;23(5):379-383. [doi: 10.1136/emj.2005.029793] [Medline: 16627841]
122. Bein KJ, Berendsen Russell S, Ní Bhraonáin S, Seimon RV, Dinh MM. Does volume or occupancy influence emergency access block? A multivariate time series analysis from a single emergency department in Sydney, Australia during the COVID‐19 pandemic. Emerg Med Australas 2021 Apr;33(2):343-348 [FREE Full text] [doi: 10.1111/1742-6723.13717]
123. Jobé J, Donneau AF, Scholtes B, Ghuysen A. Quantifying emergency department crowding: Comparison between two scores. Acta Clin Belg 2018 Jun;73(3):207-212. [doi: 10.1080/17843286.2017.1410605] [Medline: 29207925]
124. Grant KL, Bayley CJ, Premji Z, Lang E, Innes G. Throughput interventions to reduce emergency department crowding: a systematic review. CJEM 2020 Nov;22(6):864-874. [doi: 10.1017/cem.2020.426] [Medline: 33448916]
125. Soldatenkova A, Calabrese A, Levialdi Ghiron N, Tiburzi L. Emergency department performance assessment using administrative data: a managerial framework. PLoS ONE 2023;18(11):e0293401. [doi: 10.1371/journal.pone.0293401] [Medline: 37917787]
126. Kenny JF, Chang BC, Hemmert KC. Factors affecting emergency department crowding. Emerg Med Clin North Am 2020 Aug;38(3):573-587. [doi: 10.1016/j.emc.2020.04.001] [Medline: 32616280]
127. Parker BT, Marco C. Emergency department length of stay: Accuracy of patient estimates. West J Emerg Med 2014 Mar;15(2):170-175. [doi: 10.5811/westjem.2013.9.15816] [Medline: 24672606]




XSL•FO
RenderX

128. Almasi S, Rabiei R, Moghaddasi H, Vahidi-Asl M. Emergency department quality dashboard; a systematic review of performance indicators, functionalities, and challenges. Arch Acad Emerg Med 2021;9(1):e47. [doi: 10.22037/aaem.v9i1.1230] [Medline: 34405145]
129. Cho M, Song M, Park J, Yeom SR, Wang IJ, Choi BK. Process mining-supported emergency room process performance indicators. Int J Environ Res Public Health 2020 Aug 28;17(17):6290. [doi: 10.3390/ijerph17176290] [Medline: 32872350]

## Abbreviations

**ED:** emergency department
**ED LOS:** emergency department length of stay
**LWBS:** left without being seen
**OR:** odds ratio
**PRISMA-ScR:** Preferred Reporting Items for Systematic Review and Meta-Analyses extension for Scoping Reviews

*Edited by N Cahill; submitted 29.May.2025; peer-reviewed by CC Tan, E Bai, M Stoner, R Hawari; revised version received 03.Dec.2025; accepted 03.Dec.2025; published 05.May.2026.*

*Please cite as:*
*Mohd Nasir NN, Indera Putera KAS, Noris NJ, Libasin Z, Krishnan M, Salaton NF, Lum KY, Abdullah NNR, Saimy IS*
*Overcrowding Indicators in Emergency Departments Across Countries: Scoping Review*
*Interact J Med Res 2026;15:e78073*
*URL: https://www.i-jmr.org/2026/1/e78073*
*doi:10.2196/78073*

© Natasya Nur Mohd Nasir, Ku Anis Shazura Indera Putera, Nur Jihan Noris, Zalina Libasin, Muniamal Krishnan, Nor Fauziah Salaton, Kah Yee Lum, Nur Nadia Renu Abdullah, Intan Syafinaz Saimy. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 5.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Information and Communication Technologies for Chronic Disease Self-Management in Adults Aged 65 Years and Older: Scoping Review

Paul Murdock[1], MD, MS; Yiyi Wu[2], MA; Charles R Senteio[2], MSW, MBA, PhD

[1]Cleveland Clinic, 9500 Euclid Ave, Cleveland, OH, United States

[2]Rutgers University School of Communication and Information, New Brunswick, NJ, United States

**Corresponding Author:**
Paul Murdock, MD, MS
Cleveland Clinic, 9500 Euclid Ave, Cleveland, OH, United States

## Abstract

**Background:** The increasing number of older adults living with chronic conditions has led to rapid growth in information and communication technologies (ICTs) designed to support chronic disease self-management. Although many technologies target behaviors such as medication adherence, physical activity, dietary management, and follow-up care, the breadth, characteristics, and design considerations of these tools for adults aged 65 years and older have not been comprehensively reported.

**Objective:** This scoping review aims to systematically map the existing literature describing ICTs developed to support chronic disease self-management among adults aged 65 years and older. Specifically, the review seeks to (1) identify the types of ICTs available; (2) characterize the self-management behaviors they target; and (3) examine the extent of older adults' involvement in the design, adaptation, or evaluation of these technologies.

**Methods:** This review followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines. Seven databases (PubMed, CINAHL, Web of Science, Cochrane Library, Compendex, IEEE Xplore, and Computers & Applied Sciences Complete) were searched, with all searches completed on December 15, 2024. Inclusion criteria were peer-reviewed studies published in English between 2007 and 2025 that (1) included adults aged ≥65 years; (2) addressed one or more chronic diseases; and (3) evaluated, described, or tested an ICT intended to support at least 1 chronic disease self-management behavior. Two reviewers independently screened all titles and abstracts and full texts; disagreements were resolved by a third reviewer. Data were charted using a standardized extraction template and synthesized narratively by technology type and self-management domain.

**Results:** Nineteen studies met the inclusion criteria. Technologies were grouped into 4 broad categories: mobile apps, online platforms, wearable or sensor-based tools, and smart home or device-integrated systems. Physical activity and medication management were the most targeted self-management behaviors, whereas follow-up appointment adherence and dietary behaviors were less frequently addressed. Only a small number of studies explicitly involved older adults in the design or development process, and such involvement was often limited to usability testing rather than participatory co-design.

**Conclusions:** The current evidence base is fragmented, with substantial variability in technology types, targeted behaviors, and reported outcomes. Significant gaps remain regarding the participatory design of ICTs with older adults and the development of technologies that address multiple self-management needs simultaneously. Future ICT development should intentionally incorporate older adults and caregivers throughout the design cycle and expand beyond single-behavior interventions to reflect the multimorbidity common in this population.

***(Interact J Med Res 2026;15:e60542)*** doi:10.2196/60542

**KEYWORDS**

health technology; self-care; self-management; chronic disease; efficacy; technology; older adult; gerontology; chronic; chronic diseases; systematic review; behavioral health; health equity; Preferred Reporting Items for Systematic reviews and Meta-Analyses; PRISMA; medication behavior; dietary behavior; physical activity; physical rehabilitation; digital health

## Introduction

Most population projections estimate that by 2030, 1 in 6 people worldwide will be aged 60 years or older [1], compared with 1 in 10 at present [2]. The older adult population is expected to grow and eventually double by 2050, reaching 2.1 billion persons worldwide [1]. According to the US Centers for Disease Control and Prevention, 6 in 10 adults in the United States are living with 1 or more of the "big five" chronic conditions: diabetes mellitus, cardiovascular disease, chronic respiratory



XSL•FO
RenderX

disease, cancer, and stroke [3]. Older adults are at increased risk of having chronic conditions; two-thirds of Medicare beneficiaries have 2 or more chronic conditions [3]. This shifting demographic and the prevalence of chronic conditions result in an expanding number of older adults living with chronic conditions.

For individuals living with chronic diseases, effective self-management, wherein patients take an active role in their own care, is essential for improving physical health, emotional well-being, and overall quality of life [4]. Self-management typically involves patients consistently engaging in healthy lifestyle behaviors, including maintaining a balanced diet, participating in regular physical activity, adhering to prescribed medications, and attending follow-up medical appointments [5].

Rapid advancements in technology (ie, smartphones and applications), connectivity (ie, mobile broadband availability), and commercial potential have resulted in an explosion of numerous information and communications technology (ICT) tools designed to support chronic disease self-management behaviors [6-8]. For older adults in particular, technologies designed to support health care and chronic disease management are classified into four general areas, which are based upon location of use and platform: (1) mobile-based apps, (2) smart home–based technologies, (3) online-based technologies, and (4) personalized application-based technologies [9]. Mobile-based apps, typically accessed via smartphones and tablets, are widely used to support remote medical services and personalized care. Among these, mobile health and telehealth platforms are the most prevalent. Smart home technologies, which incorporate ICT-enabled tools within the home, such as digital reminders for medical appointments, are increasingly recognized for their role in chronic disease management. Online-based technologies mainly refer to web-based services, including access to podcasts, disease-specific forums, health care providers and product reviews, and other health-related content, all of which support informed decision-making and continuous engagement with care. Personalized application-based technologies refer to a wide range of assistive devices that are programmed to improve care outcomes for older adults. Examples include intelligent devices that monitor vital signs such as blood pressure.

However, using ICTs to support chronic disease self-management among older adults presents significant challenges due to age-related barriers to technology use and engagement [10]. These barriers include generally lower levels of digital literacy and skills; increased concerns about privacy; and physical or cognitive limitations such as impaired vision, hearing loss, and memory decline [11-14]. As a result, older adults often depend on caregivers to effectively use health technologies [15]. In response, the literature has consistently emphasized the need for more accessible and inclusive design, particularly for ICTs aimed at this population [16]. A widely endorsed approach is to involve both older adults and their caregivers in the design process [17,18]. This co-design strategy has been shown to enhance usability, increase patient satisfaction, improve disease management, boost health literacy, and reduce health care costs [18].

Despite substantial expansion of ICTs for chronic disease management, our search found that the literature specifically focused on adults aged 65 years and older remains fragmented and inconsistent. Prior reviews often examine younger populations, focus narrowly on single diseases (eg, diabetes), or aggregate technologies without distinguishing specific self-management behaviors. To date, no comprehensive review has mapped the breadth of ICTs supporting self-management behaviors uniquely for adults aged 65 years and older.

A scoping review is methodologically suited to (1) map broad and diverse bodies of literature, (2) clarify key concepts, (3) identify types of evidence, and (4) highlight gaps rather than evaluate intervention effectiveness or pool outcomes. Given the wide variation in study designs, technologies, and outcomes and the exploratory nature of the research questions, a scoping review is justified and aligns with PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidance.

Guided by this rationale, this scoping review aims to address the following research questions:

1. What types of ICTs have been developed to support chronic disease self-management among adults aged 65 years and older?
2. Which self-management behaviors (eg, medication adherence, physical activity, dietary management, and follow-up care) are targeted by these technologies?
3. To what extent are older adults involved in the design, adaptation, or evaluation of these technologies?

## Methods

### Protocol and Registration

This review followed the PRISMA-ScR guidelines. The protocol was not registered.

### Eligibility Criteria

Consistent with scoping review methodology, eligibility criteria were developed to capture the breadth of literature describing ICTs that support chronic disease self-management among adults aged 65 years and older, rather than to evaluate intervention effectiveness.

The inclusion and exclusion criteria are summarized in Textbox 1.



XSL•FO
RenderX

**Textbox 1.** Inclusion and exclusion criteria.

**Inclusion criteria**

- Population—adults aged 65 years or older managing 1 or more chronic diseases
- Concept—information and communication technologies (ICTs) designed to support at least 1 chronic disease self-management behavior**,** including (but not limited to) medication adherence, physical activity, diet, or follow-up appointment attendance
- Context—any setting (eg, home, community, and clinical)
- Types of sources: peer-reviewed empirical studies (qualitative, quantitative, mixed methods, feasibility or pilot studies, observational studies, or trials) published in English between 2007 and 2025

**Exclusion criteria**

- Interventions targeting only health care providers
- Nondigital or nonpersonalized technologies
- Studies not focused on chronic disease self-management
- Reviews, editorials, protocols, conference abstracts, or dissertations

A pilot screening phase was conducted during the initial article selection, in which similar articles from target journals were reviewed to refine the scope and thematic alignment of this review. This process guided database selection and search strategy refinement but was not part of the formal eligibility criteria. Because the goal of a scoping review is to map existing evidence rather than restrict it based on study design or comparator groups, no comparator was required or used as part of the eligibility criteria.

## Information Sources

A comprehensive search was conducted in 7 databases selected in collaboration with a health sciences librarian: PubMed, CINAHL, Web of Science, Cochrane Library, Compendex, IEEE Xplore, and Computers & Applied Sciences Complete. All searches were completed on December 15, 2024. Reference lists of included studies were hand searched for additional relevant articles.

## Search Strategy

Search terms were developed to reflect the review's core aims: older adults, digital health technologies, and chronic disease self-management. Terms were adapted for each database using controlled vocabulary (eg, Medical Subject Headings [MeSH]) and keywords. Detailed search strings are included in Multimedia Appendix 1.

## Study Selection

Titles and abstracts were independently screened by 2 reviewers (PM and CRS). Rayyan, an internet-based software package, was used to facilitate article screening [19]. The 2 authors independently completed the title and abstract screening and full-text screening. Full texts of potentially eligible studies were assessed using the defined inclusion and exclusion criteria. Disagreements were resolved through consensus with the third author (YW). A PRISMA-ScR flow diagram was used to outline the study identification and selection process.

## Data Collection Process

In accordance with scoping review methodology, a standardized data charting form was developed and iteratively refined. Two reviewers (PM and CRS) independently extracted data from the included studies using a standardized data extraction form. Extracted data included study characteristics (author, year, and country), population demographics, technology description, targeted chronic diseases, self-management domain, outcomes related to technology use and acceptance, and reported effectiveness. The third reviewer (YW) resolved disagreements.

## Data Items

Key data items included the following:

- Author, year, and country
- Study design
- Participant age and health status
- Type and functionality of technology
- Chronic diseases addressed
- Self-management behaviors targeted
- Setting (home based, clinical, and other)
- Extent and type of older adult involvement in design or evaluation
- Reported outcomes (clinical, behavioral, or usability related)

Data charting was iterative; reviewers updated the form as familiarity with the literature increased, in accordance with best practices for scoping reviews.

## Synthesis of Results

Consistent with PRISMA-ScR, no risk of bias or formal quality assessment was conducted, as the goal was to describe the extent and nature of existing evidence rather than to evaluate or compare intervention effectiveness.

Because of substantial heterogeneity in study designs, outcomes, technologies, and measurement approaches, meta-analysis was neither planned nor appropriate. This aligns with the purpose of a scoping review.

## Narrative and Thematic Synthesis

A descriptive, narrative synthesis was conducted. Studies were grouped according to the following categories:

- Technology type (mobile apps, web-based platforms, wearable or sensor technologies, and smart home systems)



XSL•FO
RenderX

- Targeted self-management behaviors
- Chronic disease or health focus
- Degree of older adult involvement in design or testing

To identify key themes across studies, we used a structured thematic analysis process. First, during initial coding, 2 reviewers independently reviewed charted data and coded recurring concepts related to technology use, usability, self-management support, and design involvement. Second, in code consolidation, codes were compared, merged, and refined through discussion. Third, we used theme development to group codes into preliminary categories and iteratively refined them to generate overarching themes reflecting patterns across studies. Fourth, in a final synthesis, themes were summarized and integrated into the narrative results. This analytic process enabled identification of major trends, gaps, and characteristics of the literature.

### Effect Measures

Because this is a scoping review, no standardized effect measures were calculated. When available, quantitative outcomes (eg, percentages, use statistics, and self-reported behavior changes) were summarized descriptively based on the information reported in each study.

## Results

### Study Selection

The database search identified 897 records. After deduplication, 815 (90.9%) titles and abstracts were screened. Following full-text review, 19 (2.1%) studies met the inclusion criteria. The study identification and selection process is summarized in Figure 1 (the PRISMA-ScR flow diagram).



XSL•FO
RenderX

**Figure 1.** PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flow diagram illustrating the identification, screening, eligibility assessment, and inclusion of sources of evidence.



XSL•FO
RenderX

## Characteristics of Included Studies

The 19 included studies were published between 2007 and 2025, with the majority (n=18, 94.7%) published after 2012. Studies were conducted across multiple disciplines, including rehabilitation, geriatrics, computer engineering, nursing, and digital health.

Sample sizes varied substantially, ranging from 10 to 803 participants, with most studies enrolling fewer than 120 participants. All studies included older adults aged 65 years and older, although several recruited broader adult populations and reported subgroup data for older adults.

Technologies described in the included studies (n=19) encompassed 4 primary categories: mobile apps and tablets (n=9, 47.4%), web-based platforms (n=5, 26.3%), wearable or sensor-based technologies (n=3, 15.8%), and smart home or device-integrated systems (n=2, 10.5%).

Self-management behaviors targeted most frequently included physical activity (n=11, 57.9%) and medication management (n=10, 52.6%). Dietary behavior (n=7, 36.8%) and follow-up appointment support (n=3, 15.8%) were less commonly addressed.

A detailed summary of charted study characteristics is provided in Table 1.

**Table .** Charted characteristics of included studies (n=19).

| Study | Technology type | Self-Management focus | Age (years) | Sample size, n | Notes (major study features) |
|---|---|---|---|---|---|
| Nischelwitzer [20] 2007 | Mobile medical app | Physiological tracking | 36 - 84 | 15 | Early mobile prototype; pilot feasibility |
| Ali et al [21] 2012 | Touchscreen platform | Nutritional education | 60 - 76 | 31 | Education-focused dietary interface |
| Ammann et al [22] 2012 | Web platform | Personalized physical activity | 19 - 89 | 803 | Large sample; tailored content |
| Hess et al [23] 2012 | SMS text messaging | Activity and dietary support | 40 - 69 | 47 | Combined prompting system |
| Reeder et al [24] 2013 | Digital pill dispenser | Medication management | Average 80 | 96 | Device-integrated system |
| Jiménez-Fernández et al [25] 2013 | Wireless sensors | Physiological tracking | Average 65 | 22 | Wearable monitoring |
| Ellis et al [26] 2013 | Pedometer and computer | Physical activity | Average 65.6 | 20 | Exercise-focused |
| Mira et al [27] 2014 | Tablet app | Medication management | ≥65 | 99 | Self-medication support |
| Dasgupta et al [28] 2016 | Tablet app | Activity and medication management | Average 66.8 | 16 | Multifunction platform |
| Costa et al [29] 2017 | Smart TV system | Social and health services | ≥65 | 62 | TV-based interface |
| Georgsson and Staggers [30] 2017 | SMS service | Follow-up, activity, and medication | Majority aged 60 - 69 | 10 | Multibehavior SMS support |
| Yan and Or [31] 2017 | Tablet with blood pressure and glucose meters | Physiological tracking | Average 69.9 | 119 | Integrated monitoring |
| Pariser et al [32] 2019 | Telemedicine | Access to resources | Average 66 | 76 | Virtual care |
| Lang et al [33] 2022 | Tablet | Physiological tracking | ≥65 | 116 | Remote monitoring |
| Nambisan et al [34] 2022 | Mobile app | Physical activity and diet | 60 - 80 | 20 | Behavior logging |
| Traviss-Turner et al [35] 2024 | Web program | Dietary behavior | 25 to >65 | 22 | Binge eating reduction |
| Shi et al [36] 2024 | Web program | Exercise and diet monitoring | Average 67.2 | 64 | Lifestyle modification |
| Valdeverona et al [37] 2024 | SMS texting | Diet, exercise, and medications | 37 - 82 | 38 | Multidomain support |
| Spinean et al [38] 2025 | Mobile app | Physical activity and diet | 18 to >65 | 147 | Large mixed-age sample |



XSL•FO
RenderX

## Mapping of Self-Management Behaviors

To describe the scope of self-management behaviors addressed by ICTs, the 19 included studies were categorized according to the primary behavior targeted.

Physical activity was addressed in 11 (57.9%) studies. These studies evaluated pedometers, activity monitoring mobile apps, wearable sensors, or web-based activity programs. Approaches included real-time step tracking, motivational prompts, and personalized exercise recommendations.

Medication management was targeted in 10 (52.6%) studies. Technologies included digital pillboxes, reminder systems, medication self-management apps, and SMS text messaging medication prompts. Most outcomes were descriptive, reporting perceived usefulness, adherence trends, or frequency of system use.

Dietary behavior was addressed in 7 (36.8%) studies. These ICTs focused on dietary logging, nutritional education through tablet or TV interfaces, and monitoring of self-reported dietary behaviors. Few studies provided objective dietary outcomes; most reported usability, satisfaction, or behavioral intentions.

Follow-up appointment support was addressed in 3 (15.8%) studies. Appointment reminders were delivered through SMS text messaging systems or integrated care platforms. Studies described improved perceived access to resources rather than clinical outcomes.

## Older Adult Involvement in Technology Design

Of the 19 studies, only 4 (21.1%) explicitly described involving older adults or caregivers in technology design or refinement. Of these, 2 studies used usability testing only, which occurred late in the development cycle. Only 2 studies incorporated participatory or co-design approaches, allowing older adults to contribute to early-stage feature development.

This limited involvement highlights a critical gap between design practices and the needs of older adult users—a key theme in our narrative synthesis.

A summary of user involvement is presented in Table 2.

**Table .** Extent of older adult involvement in information and communication technology design, adaptation, or evaluation (n=19).

| Study | Sample size, n | Older adult involvement during design | Type of involvement |
|---|---|---|---|
| Nischelwitzer et al [20] | 15 | No | Initial pilot testing only |
| Ali et al [21] | 31 | No | Usability testing after development |
| Ammann et al [22] | 803 | No | Not reported |
| Hess et al [23] | 47 | No | End user feedback only |
| Reeder et al [24] | 96 | No | Postdeployment evaluation |
| Jiménez-Fernández et al [25] | 22 | No | Usability evaluation |
| Ellis et al [26] | 20 | No | Usability or feasibility testing |
| Mira et al [27] | 99 | *Yes*[a] | *Co-design: medication routine insights* |
| Dasgupta et al [28] | 16 | No | Usability and performance testing |
| Costa et al [29] | 62 | No | Informal feedback |
| Georgsson and Staggers [30] | 10 | No | Satisfaction survey |
| Yan and Or [31] | 119 | No | Logging and perceived usefulness |
| Pariser et al [32] | 76 | No | Telemedicine use feedback |
| Lang et al [33] | 116 | No | Usability testing |
| Nambisan et al [34] | 20 | *Yes* | *Iterative prototyping* |
| Traviss-Turner et al [35] | 22 | No | Behavior tracking input |
| Shi et al [36] | 64 | No | Attitudes and satisfaction assessment |
| Valdeverona et al [37] | 38 | No | Acceptability evaluation |
| Spinean et al [38] | 147 | No | End user adoption data |

[a]Italicized text denotes studies with meaningful (but limited) co-design or development-stage input.

## Themes Identified in the Narrative Synthesis

Through thematic analysis of charted data, 3 overarching themes emerged.

### *Theme 1: ICTs Commonly Target Single Behaviors Rather Than Multidimensional Self-Management*

Most technologies (14/19, 73.7%) focused on only one self-management behavior, despite the high prevalence of multimorbidity in older adults. Physical activity and medication



XSL•FO
RenderX

adherence dominated the intervention landscape, while diet and follow-up behaviors were underrepresented.

### *Theme 2: Limited Integration of Older Adults in Design and Development*

Consistent with prior literature on participatory design, few studies included older adults in the design process, and involvement was often superficial. Studies that incorporated older adult or caregiver feedback reported improved usability, increased engagement, and greater perceived relevance. However, participatory co-design remains the exception rather than the norm.

### *Theme 3: Focus on Usability Over Effectiveness*

Across studies, outcomes overwhelmingly emphasized usability, perceived usefulness, intention to use, and satisfaction. Few studies measured changes in actual self-management behavior or clinical outcomes, reflecting a broader trend toward feasibility or proof of concept research rather than rigorous evaluation.

## Summary of Findings

The evidence base is diverse but fragmented. ICTs designed to support chronic disease self-management in older adults vary in purpose, technology type, and targeted behavior. Several key gaps were identified. These gaps included minimal involvement of older adults in technology design, scarcity of multidimensional or integrated self-management technologies, lack of objective outcome measures, and limited focus on follow-up appointment adherence and dietary behavior.

## Additional Analyses: Qualitative Synthesis

Articles that reported technology interventions and included self-management aimed at improving chronic disease outcomes using either clinical or behavioral outcomes were eligible for systematic review inclusion (Table 3). We categorized the interventions into the following 4 self-management activities: medication behavior, physical activity, dietary behavior, and follow-up appointment attendance.



XSL•FO
RenderX

**Table .** Self-management behaviors targeted by information and communication technologies in the included studies (n=19).

| Study | Medication behavior | Follow-up appointment attendance | Physical activity | Dietary behavior | Reported outcomes related to behavior |
|---|---|---|---|---|---|
| Nischelwitzer et al [20] | No | No | Yes | No | Measured user input data including blood pressure and glucose level |
| Ali et al [21] | No | No | No | Yes | Measured perceived usefulness and ease of use |
| Ammann et al [22] | No | No | Yes | No | —[a] |
| Hess et al [23] | Yes | No | No | No | Measured glucose readings and appointment attendance |
| Reeder et al [24] | Yes | No | No | No | Measured perceived ease of use and usefulness |
| Jiménez-Fernández et al [25] | Yes | No | No | No | Measured degree of satisfaction and perceived ease of use |
| Ellis et al [26] | No | No | Yes | No | Measured walking activity and speed |
| Mira et al [27] | Yes | No | No | No | Measured adherence and missed doses |
| Dasgupta et al [28] | Yes | No | Yes | No | Measured health management skills, risk for depression, and self-reported physical activity |
| Costa et al [29] | Yes | Yes | Yes | No | Measured user perception |
| Georgsson and Staggers [30] | Yes | Yes | Yes | Yes | Measured perceived improvement |
| Yan and Or [31] | Yes | No | No | No | Measured actual use and perceived usefulness |
| Pariser et al [32] | Yes | No | No | No | Measured perceived access to clinical resources |
| Lang et al [33] | Yes | No | No | No | Measured actual use and perceived usefulness |
| Nambisan et al [34] | No | No | Yes | Yes | Measured physical activity and dietary log; condition tracking |
| Traviss-Turner et al [35] | No | No | No | Yes | Measured binge eating rate |
| Shi et al [36] | No | No | Yes | Yes | Measured improvement in exercise and dietary behaviors |
| Valdeverona et al [37] | Yes | — | — | — | Measured perceived usefulness and frequency of use |
| Spinean et al [38] | No | No | Yes | Yes | Measured adherence to physical activity recommendation and dietary recommendation |



XSL•FO
RenderX

[a]Indicates that the applicable self-management behavior was not measured or included in the referenced study.

## Discussion

### Principal Findings

Despite the extensive body of literature on technology use in chronic disease management, relatively few studies (n=19) have explicitly examined the role of ICTs in supporting self-management behaviors among adults aged 65 years and older. This finding underscores the relative underdevelopment of an evidence base that is both age specific and behaviorally grounded, despite the high burden of multimorbidity and chronic disease management demands in this population. Given the well-established link between the 4 key self-management behaviors (ie, medication adherence, attending medical appointments, engaging in physical activity, and maintaining a healthy diet) and chronic disease outcomes, future research on technology use in chronic disease management should specifically address these behaviors to ensure more concrete and actionable findings [18].

Importantly, the scoping nature of this review allows for identification of patterns and gaps across heterogeneous study designs rather than assessment of intervention effectiveness. Viewed through this lens, the limited number of studies focused explicitly on adults aged 65 years and older reflects not only a quantitative gap but also a conceptual one in how older adults are positioned within digital health research. Future studies should place particular emphasis on adults aged 65 years and older, a population with a high prevalence of multiple chronic conditions and a documented lower intention to use health information technologies designed for self-management [39,40].

The identified gap gains further significance in light of known predictors of technology use for disease management, particularly performance expectancy and social influence. Lower levels of these factors may contribute to suboptimal adoption and sustained use of ICTs among older adults with chronic diseases [41]. While several studies implicitly acknowledged these determinants, few explicitly incorporated them into intervention design or evaluation frameworks. Understanding these predictors is crucial for informing targeted interventions that address the specific needs, expectations, and social contexts of older adults.

An increasing body of literature highlights the importance of involving older adults in the design of technologies intended to support their health [9]. However, only a small subset of studies in this review reported meaningful involvement of older adults during early design or development phases, with most limiting engagement to late-stage usability testing. Emerging research further suggests that incorporating older adults' caregivers and support networks into the design process can improve communication, strengthen social relationships, and enhance sustained technology use [17,42]. The limited adoption of participatory and co-design approaches identified in this review therefore represents a missed opportunity to align ICT development with the lived realities of aging with chronic disease.

Regarding technology types, mobile technologies and personalized applications were the most frequently reported, surpassing web-based and home-based solutions. This pattern aligns with broader literature on technology acceptance among older adults, which suggests that tablets and smartphones are often perceived as more accessible and easier to use than desktop or laptop computers [43,44]. However, the predominance of mobile solutions should not be interpreted as evidence of optimal fit for all older adults, particularly those with sensory, cognitive, or socioeconomic barriers that may limit access or sustained use.

The included studies addressed self-management behaviors related to physical activity, medication management, diet and nutrition, and follow-up appointment attendance. While social services and social connectedness were rarely targeted, 1 study using an interactive TV-based platform demonstrated promising outcomes related to access to services and social engagement, suggesting a potential role for ICTs in addressing social isolation as a component of chronic disease self-management [45,46]. This finding highlights the need to broaden conceptualizations of self-management beyond biomedical behaviors to include social and contextual dimensions that influence health and well-being in later life.

### Implications for Practice and Future Research

This scoping review reveals several important opportunities for improving the design and implementation of ICTs that support chronic disease self-management in older adults.

First, expand the range and integration of self-management behaviors addressed. Future technologies should move beyond single-behavior interventions to reflect the multidimensional nature of chronic disease in older adults. Attention should also be given to social services as a self-management behavior, and exploration of technologies targeting social isolation is warranted.

Second, integrate older adults and caregivers across all stages of technology development. Participatory and co-design methodologies should be prioritized to ensure that ICTs align with users' functional abilities, digital literacy, and everyday care contexts. Early-stage involvement, rather than post hoc usability testing alone, is particularly important for improving relevance and adoption.

Third, incorporate social and contextual determinants of technology use. Factors such as digital literacy, socioeconomic status, internet access, and social support networks are central to understanding ICT adoption and effectiveness among older adults. Future research should explicitly measure and report these contextual factors rather than treating them as background characteristics [43,47].

Fourth, adopt more robust evaluation approaches as technologies mature. While feasibility and usability studies remain appropriate at early stages, later-phase research should incorporate objective measures of behavior change, care processes, or health outcomes to better assess real-world impact.



XSL•FO
RenderX

Finally, develop technologies that support care continuity and follow-up. Appointment adherence and communication with health care providers remain underexplored yet highly relevant domains for older adults' health outcomes and health care use.

Collectively, these implications underscore the need for more holistic, user-centered, and contextually informed ICT development strategies tailored to older adults managing chronic disease.

### Strengths and Limitations

This review's strengths include a comprehensive, librarian-assisted search strategy; adherence to PRISMA-ScR reporting guidelines; and a structured, iterative approach to data charting and thematic synthesis. By focusing on self-management behaviors rather than specific diseases or technologies alone, this review offers a behaviorally grounded map of the current evidence base that complements prior disease-specific reviews.

Several limitations should be noted. First, while this scoping review provides valuable insights into ICTs targeting older adults, it did not assess intervention effectiveness or quality, consistent with scoping review methodology. Second, the heterogeneity of study designs, populations, and outcome measures limited cross-study comparability. Third, gray literature was not included, which may have excluded emerging or non–peer-reviewed technologies relevant to older adults. Finally, although sociocultural and structural factors were identified as important gaps, these were infrequently reported in the included studies, limiting deeper analysis of their influence on ICT adoption and use [46,48].

## Conclusions

This scoping review maps the current landscape of ICTs designed to support chronic disease self-management among adults aged 65 years and older, revealing a fragmented and uneven evidence base. Although a wide range of technologies has been developed, most focus on single self-management behaviors and provide limited evidence of meaningful older adult involvement in design or development.

The findings have important implications for aging researchers, health informatics scholars, and technology developers. Addressing the identified gaps, particularly the limited use of participatory design, the narrow focus on individual behaviors, and the lack of attention to social and contextual factors, will be essential for advancing more inclusive and effective ICT-based self-management support for older adults. As the population ages and the prevalence of multimorbidity increases, intentional, age-centered approaches to digital health design will be critical for improving chronic disease management and health equity in later life.

## Funding

This study received no external funding.

## Authors' Contributions

All 3 authors contributed equally to the development of the manuscript. Each author substantially participated in (1) the conception and design of the study, data collection, data analysis, and interpretation; (2) drafting and revising the manuscript for important intellectual content; and (3) approving the final version of the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Databases and search strings.
[DOCX File, 18 KB - i-jmr_v15i1e60542_app1.docx ]

Checklist 1
PRISMA checklist.
[DOCX File, 22 KB - i-jmr_v15i1e60542_app2.docx ]

## References

1. Ageing and health. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2023-02-18]
2. World population by age and region 2022. Statista. URL: https://www.statista.com/statistics/265759/world-population-by-age-and-region/ [accessed 2023-12-05]
3. Salive ME. Multimorbidity in older adults. Epidemiol Rev 2013;35(1):75-83. [doi: 10.1093/epirev/mxs009] [Medline: 23372025]
4. Grady PA, Gough LL. Self-management: a comprehensive approach to management of chronic conditions. Am J Public Health 2014 Aug;104(8):e25-e31. [doi: 10.2105/AJPH.2014.302041] [Medline: 24922170]



XSL•FO
RenderX

5. Allegrante JP, Wells MT, Peterson JC. Interventions to support behavioral self-management of chronic diseases. Annu Rev Public Health 2019 Apr 1;40:127-146. [doi: 10.1146/annurev-publhealth-040218-044008] [Medline: 30601717]
6. Celler BG, Lovell NH, Basilakis J. Using information technology to improve the management of chronic disease. Med J Aust 2003 Sep 1;179(5):242-246. [doi: 10.5694/j.1326-5377.2003.tb05529.x] [Medline: 12924970]
7. El-Gayar O, Timsina P, Nawar N, Eid W. Mobile applications for diabetes self-management: status and potential. J Diabetes Sci Technol 2013 Jan 1;7(1):247-262. [doi: 10.1177/193229681300700130] [Medline: 23439183]
8. Ye Q, Boren SA, Khan U, Kim MS. Evaluation of functionality and usability on diabetes mobile applications: a systematic literature review. In: Duffy V, editor. Digital Human Modeling: Springer; 2017. [doi: 10.1007/978-3-319-58466-9_11]
9. Miah SJ. Information and communication technology-based innovations for aging healthcare: a literature review. Int J Comput Healthc 2014;2(1):28-42. [doi: 10.1504/IJCIH.2014.065810]
10. Bohingamu Mudiyanselage S, Stevens J, Watts JJ, et al. Personalised telehealth intervention for chronic disease management: a pilot randomised controlled trial. J Telemed Telecare 2019 Jul;25(6):343-352. [doi: 10.1177/1357633X18775850] [Medline: 29793387]
11. Quan-Haase A, Martin K, Schreurs K. Interviews with digital seniors: ICT use in the context of everyday life. Inf Commun Soc 2016;19(5):691-707. [doi: 10.1080/1369118X.2016.1140217]
12. Fischer SH, David D, Crotty BH, Dierks M, Safran C. Acceptance and use of health information technology by community-dwelling elders. Int J Med Inform 2014 Sep;83(9):624-635. [doi: 10.1016/j.ijmedinf.2014.06.005] [Medline: 24996581]
13. Khosravi P, Ghapanchi AH. Investigating the effectiveness of technologies applied to assist seniors: a systematic literature review. Int J Med Inform 2016 Jan;85(1):17-26. [doi: 10.1016/j.ijmedinf.2015.05.014] [Medline: 26216463]
14. Mitzner TL, Sanford JA, Rogers WA. Closing the capacity-ability gap: using technology to support aging with disability. Innov Aging 2018;2(1):igy008. [doi: 10.1093/geroni/igy008] [Medline: 30480132]
15. Finkelstein R, Wu Y, Brennan-Ing M. Older adults' experiences with using information and communication technology and tech support services in New York City: findings and recommendations for post-pandemic digital pedagogy for older adults. Front Psychol 2023 Apr 17;14:1129512. [doi: 10.3389/fpsyg.2023.1129512] [Medline: 37138998]
16. Stamate A, Marzan MD, Velciu M, Paul C, Spiru L. Advancing user-centric design and technology adoption for aging populations: a multifaceted approach. Front Public Health 2024 Dec 6;12:1469815. [doi: 10.3389/fpubh.2024.1469815] [Medline: 39712308]
17. Swallow D, Petrie H, Power C, Lewis A, Edwards AD. Involving older adults in the technology design process: a case study on mobility and wellbeing in the built environment. Stud Health Technol Inform 2016;229:615-623. [Medline: 27534357]
18. Fischer B, Peine A, Östlund B. The importance of user involvement: a systematic review of involving older users in technology design. Gerontologist 2020 Oct;60(7):e513-e523. [doi: 10.1093/geront/gnz163]
19. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5:210. [doi: 10.1186/s13643-016-0384-4]
20. Nischelwitzer A, Pintoffl, K, Loss C, Holzinger A. Design and development of a mobile medical application for the management of chronic diseases: methods of improved data input for older people. In: Holzinger A, editor. HCI and Usability for Medicine and Health Care: Springer; 2007. URL: http://link.springer.com/10.1007/978-3-540-76805-0 [doi: 10.1007/978-3-540-76805-0_10]
21. Ali NM, Shahar S, Kee YL, Norizan AR, Noah SAM. Design of an interactive digital nutritional education package for elderly people. Inform Health Soc Care 2012 Dec;37(4):217-229. [doi: 10.3109/17538157.2012.654843] [Medline: 22583111]
22. Ammann R, Vandelanotte C, de Vries H, Mummery WK. Can a website-delivered computer-tailored physical activity intervention be acceptable, usable, and effective for older people? Health Educ Behav 2013 Apr;40(2):160-170. [doi: 10.1177/1090198112461791] [Medline: 23077157]
23. Hess R, Fischer G, Weimer M. Abstracts from the 35th Annual Meeting of the Society of General Internal Medicine [Intensity of messaging necessary to encourage patients to access the PHR: preliminary results from the Smart-PHR study]. J Gen Intern Med 2012;Suppl 2(Suppl 2):S231. [doi: 10.1007/s11606-012-2038-0] [Medline: 22528616]
24. Reeder B, Demiris G, Marek KD. Older adults' satisfaction with a medication dispensing device in home care. Inform Health Soc Care 2013 Sep;38(3):211-222. [doi: 10.3109/17538157.2012.741084] [Medline: 23323721]
25. Jiménez-Fernández S, de Toledo P, del Pozo F. Usability and interoperability in wireless sensor networks for patient telemonitoring in chronic disease management. IEEE Trans Biomed Eng 2013 Dec;60(12):3331-3339. [doi: 10.1109/TBME.2013.2280967] [Medline: 24021636]
26. Ellis T, Latham NK, DeAngelis TR, Thomas CA, Saint-Hilaire M, Bickmore TW. Feasibility of a virtual exercise coach to promote walking in community-dwelling persons with Parkinson disease. Am J Phys Med Rehabil 2013 Jun;92(6):472-481. [doi: 10.1097/PHM.0b013e31828cd466] [Medline: 23552335]
27. Mira JJ, Navarro I, Botella F, et al. A Spanish pillbox app for elderly patients taking multiple medications: randomized controlled trial. J Med Internet Res 2014 Apr 4;16(4):e99. [doi: 10.2196/jmir.3269] [Medline: 24705022]




XSL•FO
RenderX

28. Dasgupta D, Reeves KG, Chaudhry B, Duarte M, Chawla NV. ESeniorCare: technology for promoting well-being of older adults in independent living facilities. Presented at: 2016 IEEE International Conference on Healthcare Informatics (ICHI). [doi: 10.1109/ICHI.2016.86]
29. Costa CR, Anido-Rifon LE, Fernandez-Iglesias MJ. An open architecture to support social and health services in a smart TV environment. IEEE J Biomed Health Inform 2017 Mar;21(2):549-560. [doi: 10.1109/JBHI.2016.2525725] [Medline: 26863683]
30. Georgsson M, Staggers N. Patients' perceptions and experiences of a mHealth diabetes self-management system. Comput Inform Nurs 2017 Mar;35(3):122-130. [doi: 10.1097/CIN.0000000000000296] [Medline: 27748662]
31. Yan M, Or C. A 12-week pilot study of acceptance of a computer-based chronic disease self-monitoring system among patients with type 2 diabetes mellitus and/or hypertension. Health Informatics J 2019 Sep;25(3):828-843. [doi: 10.1177/1460458217724580] [Medline: 28820007]
32. Pariser P, Pham TNT, Brown JB, Stewart M, Charles J. Connecting people with multimorbidity to interprofessional teams using telemedicine. Ann Fam Med 2019 Aug 12;17(Suppl 1):S57-S62. [doi: 10.1370/afm.2379] [Medline: 31405877]
33. Lang C, Voigt K, Neumann R, Bergmann A, Holthoff-Detto V. Adherence and acceptance of a home-based telemonitoring application used by multi-morbid patients aged 65 years and older. J Telemed Telecare 2022 Jan;28(1):37-51. [doi: 10.1177/1357633X20901400] [Medline: 32009577]
34. Nambisan P, Stange KC, Lyytinen K, Kahana E, Duthie E, Potnek M. A comprehensive digital self-care support system for older adults with multiple chronic conditions: development, feasibility, and usability testing of myHESTIA. J Appl Gerontol 2023 Feb;42(2):170-184. [doi: 10.1177/07334648221129859] [Medline: 36226748]
35. Traviss-Turner G, Bowes E, Hill A, et al. Providing online guided Self-help for the management of binge eating in adults with type 2 diabetes: the POSE-D pilot study and process evaluation. Br J Nutr 2024 Dec 14;132(11):1542-1552. [doi: 10.1017/S0007114524002599] [Medline: 39526432]
36. Shi Y, Stanmore E, McGarrigle L, et al. Development of a community intervention combining social media-based health education plus exercise programme (SHEEP) to improve muscle function among young-old adults with possible sarcopenia: co-design approach. Maturitas 2024 Aug;186:108027. [doi: 10.1016/j.maturitas.2024.108027] [Medline: 38820935]
37. Valdeverona K, Hughes L, Savory MA, Padilla BI. The use of SMS text messaging to improve glycemic control, diabetes self-management, and patient satisfaction. J Nurse Pract 2024 May;20(5):104956. [doi: 10.1016/j.nurpra.2024.104956]
38. Spinean A, Mladin A, Carniciu S, Stănescu AMA, Serafinceanu C. Emerging methods for integrative management of chronic diseases: utilizing mHealth apps for lifestyle interventions. Nutrients 2025 Apr 29;17(9):1506. [doi: 10.3390/nu17091506] [Medline: 40362815]
39. Cutilli CC, Simko LC, Colbert AM, Bennett IM. Health literacy, health disparities, and sources of health information in U.S. older adults. Orthop Nurs 2018;37(1):54-65. [doi: 10.1097/NOR.0000000000000418] [Medline: 29369135]
40. Fernández-Ballesteros R, Molina M, Schettini R, del Rey Á. Promoting active aging through university programs for older adults: an evaluation study. GeroPsych (Bern) 2012;25(3):145-154. [doi: 10.1024/1662-9647/a000064]
41. Yousef CC, Salgado TM, Farooq A, et al. Predicting patients' intention to use a personal health record using an adapted unified theory of acceptance and use of technology model: secondary data analysis. JMIR Med Inform 2021 Aug 17;9(8):e30214. [doi: 10.2196/30214] [Medline: 34304150]
42. Dykes S, Chu CH. Now more than ever, nurses need to be involved in technology design: lessons from the COVID-19 pandemic. J Clin Nurs 2021 Apr;30(7-8):e25-e28. [doi: 10.1111/jocn.15581] [Medline: 33289230]
43. Damant J, Knapp M. What are the likely changes in society and technology which will impact upon the ability of older adults to maintain social (extra-familial) networks of support now, in 2025 and in 2040? : Government Office for Science, UK; 2015 URL: https://researchonline.lse.ac.uk/id/eprint/63146/ [accessed 2026-02-23]
44. Wilson SA, Byrne P, Rodgers SE, Maden M. A systematic review of smartphone and tablet use by older adults with and without cognitive impairment. Innov Aging 2022 Jan 6;6(2):igac002. [doi: 10.1093/geroni/igac002] [Medline: 35243008]
45. Boulos C, Salameh P, Barberger-Gateau P. Social isolation and risk for malnutrition among older people. Geriatr Gerontol Int 2017 Feb;17(2):286-294. [doi: 10.1111/ggi.12711] [Medline: 26790629]
46. Tomaka J, Thompson S, Palacios R. The relation of social isolation, loneliness, and social support to disease outcomes among the elderly. J Aging Health 2006 Jun;18(3):359-384. [doi: 10.1177/0898264305280993] [Medline: 16648391]
47. Yu RP, Ellison NB, McCammon RJ, Langa K. Mapping the two levels of digital divide: internet access and social network site adoption among older adults in the USA. Inf Commun Soc 2015;19(10):1-20. [doi: 10.1080/1369118X.2015.1109695]
48. Diamantidis CJ, Becker S. Health information technology (IT) to improve the care of patients with chronic kidney disease (CKD). BMC Nephrol 2014 Jan 9;15:7. [doi: 10.1186/1471-2369-15-7] [Medline: 24405907]


## Abbreviations

**ICT:** information and communication technology
**MeSH:** Medical Subject Headings
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews



XSL•FO
RenderX

*Edited by M Balcarras, N Cahill; submitted 14.May.2024; peer-reviewed by F Bastos, ME Gagnon, U Sinha; revised version received 13.Jan.2026; accepted 13.Jan.2026; published 19.Mar.2026.*

*Please cite as:*
*Murdock P, Wu Y, Senteio CR*
*Information and Communication Technologies for Chronic Disease Self-Management in Adults Aged 65 Years and Older: Scoping Review*
*Interact J Med Res 2026;15:e60542*
*URL: https://www.i-jmr.org/2026/1/e60542*
*doi:10.2196/60542*

© Paul Murdock, Yiyi Wu, Charles R Senteio. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 19.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Global Research Trends and Hotspots in Gene Editing and Stem Cell Therapies for Neurodegenerative Diseases: Bibliometric and Visualization Analysis

Lijun Xiang[1*], PhD; Yun Xiao[2*], PhD; Ming Cai[3], PhD; Jing Qin[1], PhD; Ting Wang[3], PhD; Xueming Xiang[4], PhD; Jun Ke[1], PhD; Ganlin Peng[2], PhD

[1]Department of General Medicine, Zhangjiajie Hospital Affiliated to Hunan Normal University, Zhangjiajie, Hunan, China

[2]Department of Respiratory and Critical Care Medicine, Changsha Central Hospital, No. 161, Shaoshan South Road, Tianxin District, Changsha, Hunan, China

[3]Department of Neurology, Zhangjiajie Hospital Affiliated to Hunan Normal University, Zhangjiajie, Hunan, China

[4]Department of Infectious Diseases, Zhangjiajie Hospital Affiliated to Hunan Normal University, Zhangjiajie, Hunan, China

[*]these authors contributed equally

**Corresponding Author:**
Ganlin Peng, PhD
Department of Respiratory and Critical Care Medicine, Changsha Central Hospital, No. 161, Shaoshan South Road, Tianxin District, Changsha, Hunan, China

## Abstract

**Background:** Neurodegenerative diseases are a major and growing global health burden. Their pathogenesis is complex, and effective therapies remain limited. Gene editing and stem cell–based strategies are reshaping the therapeutic landscape. However, the field has not been systematically examined through bibliometric analysis.

**Objective:** We aimed to define the intellectual landscape of global research on gene editing and stem cell therapy for neurodegenerative diseases from 2005 to 2024, highlight evolving hotspots, track the field's evolution, and identify major bottlenecks limiting clinical translation.

**Methods:** We retrieved 1821 publications from the Web of Science Core Collection (2005-2024). We performed a multidimensional bibliometric analysis using CiteSpace and VOSviewer. We assessed publication output, country and institutional contributions, key authors and journals, co-cited references, and keyword networks. These analyses were used to track the field's evolution and pinpoint emerging themes.

**Results:** In total, 9978 researchers from 90 countries and 2515 institutions contributed to this literature. Annual publications increased from 28 in 2005 to 179 in 2024, with stepwise growth over time. The United States ranked first in output (n=780) and in citation impact (total local citation score=2784; total global citation score=40,009). China and India ranked second and fifth in output, respectively, but their average citation impact was lower than that of the leading countries. The University of California, San Francisco, and Johns Hopkins University remained consistently influential. Boulis NM, Bankiewicz KS, and Feldman EL were among the most prominent contributors. *Molecular Therapy* was the leading journal in this area. Keyword analyses pointed to a growing intersection between genetics and immunology. Major topics included nanotechnology-based delivery, adeno-associated virus vectors, small interfering RNA, intrathecal microsphere injection, autophagy, blood-brain barrier (BBB) targeting, clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9), and induced pluripotent stem cells. Burst detection highlighted "open label" as a recent hotspot. This likely reflects rising translational activity and early clinical testing.

**Conclusions:** The field is moving from technology development toward clinical translation. Anglo-American countries currently drive both productivity and influence. China and India contribute heavily to volume but need a stronger impact. CRISPR/induced pluripotent stem cell platforms and BBB-focused delivery remain central frontiers. The rise of "open-label" studies suggests accelerating clinical momentum. Future progress will require safer and more efficient delivery, clearer standards, and larger global consortia to harmonize protocols and speed translation.

***(Interact J Med Res 2026;15:e83709)*** doi:10.2196/83709

**KEYWORDS**

neurodegenerative diseases; gene editing technology; stem cell therapy; bibliometric analysis; clinical translation



XSL•FO
RenderX

## Introduction

Neurodegenerative diseases (NDs) comprise a group of progressive disorders characterized by neuronal loss and functional decline, leading to worsening cognitive, behavioral, and/or motor impairments over time [1]. Common NDs include Alzheimer disease (AD), Parkinson disease (PD), amyotrophic lateral sclerosis (ALS), and Huntington disease (HD). Globally, more than 50 million individuals are affected, and the prevalence is rising with population aging, creating a growing public health and economic burden [2]. Although decades of research have clarified key pathological hallmarks, such as β-amyloid deposition, tau hyperphosphorylation, and α-synuclein aggregation [3,4], most available therapies remain largely symptomatic and do not reliably halt or reverse disease progression. Only a limited number of drugs, including donepezil and levodopa, are approved for clinical use, and their benefits are modest and often constrained by adverse effects. Accordingly, there remains a strong need for therapies that modify disease mechanisms rather than symptoms alone [5,6].

Recent advances in gene editing and stem cell technologies have created new therapeutic possibilities for NDs [7,8]. The clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) system enables precise perturbation of disease-relevant genes and can be used to correct pathogenic variants or modulate aberrant protein expression, including targets such as *APP*, *SNCA*, *SOD1*, and *C9orf72*. In parallel, induced pluripotent stem cells (iPSCs) can be differentiated into functional neuronal populations, providing a platform for cell replacement strategies. For example, CRISPR-based *SOD1* silencing delays motor neuron degeneration in ALS models, and transplantation of iPSC-derived dopaminergic neurons has shown potential to improve motor outcomes in clinical studies of PD [9]. Notably, the convergence of these approaches is becoming an active research direction. The Cas9nVQR editing tool can efficiently modulate gene dosage in human iPSCs and has been reported to ameliorate AD-related phenotypes, including amyloid-β (Aβ) secretion and tau hyperphosphorylation [10]. Gene editing may also be applied to improve graft survival and functional engraftment by reducing immune rejection after transplantation [11]. However, key challenges remain, including off-target editing, tumorigenic risk of stem cell–derived products, and safe and efficient delivery in vivo. Addressing these barriers will be essential for clinical translation.

Bibliometric analysis applies quantitative methods to published literature to map the knowledge structure and research dynamics within a field. Using tools, such as CiteSpace, VOSviewer, and HistCite, bibliometrics can identify influential contributors and core documents, track emerging topics, and evaluate scholarly impact. It can also reveal collaboration networks and cross-disciplinary opportunities, thereby informing more strategic allocation of research efforts and resources. In addition, bibliometric evidence can help policymakers understand field trajectories and support evidence-based decision-making [12-14]. In NDs, prior bibliometric studies have focused on selected diagnostic and therapeutic topics [15,16]. However, gene editing and stem cell therapy represent rapidly expanding and increasingly intersecting frontiers, and a systematic bibliometric assessment of their progress, technical barriers, and translational trajectories in NDs remains lacking. Therefore, we performed a bibliometric and visual analysis of global research on gene editing and stem cell therapy for NDs from 2005 to 2024. By integrating publication trends, institutional collaboration networks, co-citation clustering, and keyword burst detection, this study aims to define the field's intellectual landscape, highlight evolving hotspots, and identify major bottlenecks limiting clinical translation [17].

## Methods

### Data Sources and Search Strategies

This study systematically retrieved articles published between January 2005 and December 2024 based on the Web of Science Core Collection (WoSCC) database. The retrieval strategy revolved around 2 major themes: NDs (including AD, PD, HD, etc) and gene editing and stem cell therapy technologies (covering CRISPR/Cas9, base editing, stem cell transplantation, etc). The 2 groups of terms were connected by the Boolean operator "AND." For the ND group, an extended search formula was used as follows: TS=("Neurodegenerative Diseases" OR "Degenerative Neurologic Disorder" OR "Nervous System Degenerative Diseases" OR "Neurodegenerative Disorder" OR "Neurologic Degenerative Disease" OR "Neurologic Degenerative Condition" OR "Neurologic Degenerative Diseases" OR "Degenerative Neurologic Disease" OR "Alzheimer Disease" OR "Alzheimer Syndrome" OR "Alzheimer Type Dementia" OR "Senile Dementia" OR "Parkinson Disease" OR "Huntington Disease" OR "Chronic Progressive Hereditary Chorea" OR "Huntington Chorea" OR "Amyotrophic Lateral Sclerosis" OR "Frontotemporal Dementia" OR "Frontotemporal Lobe Dementia" OR "Multiple System Atrophy" OR "Spinocerebellar Ataxia" OR "Prion Diseases"). The gene editing and stem cell group integrated technical keywords as follows: TS= ("Gene Editing" OR "Genome Editing" OR "Base Editing" OR "Genetic Therapy" OR "Genetic Therapies" OR "DNA Therapy" OR "Gene Therapy" OR "Gene-Based Therapy" OR "Stem Cell Therapy" OR "Stem Cell Treatment" OR "Stem Cell Transplantation"). Moreover, the search terms made full use of the truncation symbol "*" to cover term variants (such as "therapy" and "therapies"). The literature screening was carried out in 2 stages: preliminary screening and manual screening. In the preliminary screening stage, nonresearch papers/review literature (such as conference abstracts and patents) and non-English literature were first excluded. In the manual screening stage, 2 researchers independently evaluated the literature, and they mainly excluded the literature that did not match the disease type and the degenerative lesions (such as stroke and epilepsy) and the literature that did not match the treatment methods (such as literature only involving traditional drugs or nongenetic intervention methods). When there were disagreements during the screening process, a consensus was reached through discussion to ensure that the finally included literature met both the relevance of the target disease and the consistency of the treatment methods. Finally, we exported the relevant data of



XSL•FO
RenderX

the literature (authors, institutions, keywords, references, etc) in plain text format.

### Research Methods

We built a reproducible, multistep workflow and applied complementary software tools for descriptive statistics, collaboration mapping, and knowledge-structure mining. We first used HistCite (v12.03.17; developed by Eugene Garfield) to summarize annual publication output and overall growth of the field. To assess citation impact, we calculated the total local citation score (TLCS) and total global citation score (TGCS). The TLCS captures a paper's influence within the dataset, whereas the TGCS reflects its impact across the broader scientific literature. Together, these metrics provide a domain-specific and field-wide view of research influence.

Next, we used VOSviewer (v1.6.19; developed by Nees Jan van Eck and Ludo Waltman at Leiden University's Centre for Science and Technology Studies) to construct co-authorship networks at the country, institution, and author levels. To focus on core contributors, we applied minimum thresholds of 20 publications for countries and institutions and 10 publications for authors. In the resulting maps, node size is proportional to publication output, and edge thickness represents collaboration strength. Overlay visualizations encode the average publication year as a color gradient, allowing temporal shifts in activity to be viewed directly [18,19].

To characterize the intellectual structure and identify emerging fronts, we applied CiteSpace (v6.3.R1; developed by Chaomei Chen at Drexel University). We performed document co-citation clustering using 1-year time slices, selected nodes with the g-index (k=25), and pruned networks with the Pathfinder algorithm to emphasize the strongest connections. Cluster quality was evaluated using modularity (Q) and mean silhouette (S) values, with Q >0.3 indicating a meaningful modular structure and S >0.5 indicating acceptable internal consistency. All clustering analyses in this study met these criteria. We then used keyword co-occurrence and burst detection to define major themes and track rapidly rising topics. Finally, journal dual-map overlays were used to visualize cross-disciplinary citation flows [20].

## Results

### Annual Publication Trend Analysis

We identified 1821 publications (2005‐2024) (Figure 1), including 1083 original articles (59.5%) and 738 reviews (40.5%). Articles accumulated a TLCS of 3343 and a TGCS of 45,163, whereas reviews had a TLCS of 1443 and a TGCS of 34,729. The predominance of original articles is consistent with a technology-driven surge in primary discovery, while the substantial share of reviews suggests that multiple subareas have begun to consolidate into more mature themes. Citation patterns differed by publication type. Articles showed a lower TGCS-to-TLCS ratio than reviews (13.5 vs 24.1), indicating that original studies tended to concentrate influence within the core literature of this domain, while reviews more readily propagated across adjacent fields. Together, these data suggest a field still fueled by primary innovation but increasingly supported by integrative synthesis.



XSL•FO
RenderX

**Figure 1.** Flowchart of data collection in this study. The diagram illustrates the identification, screening, eligibility, and inclusion phases for records retrieved from the Web of Science Core Collection. The numbers of records identified, included, and excluded at each stage are shown, with specific reasons for exclusion provided in the corresponding boxes.

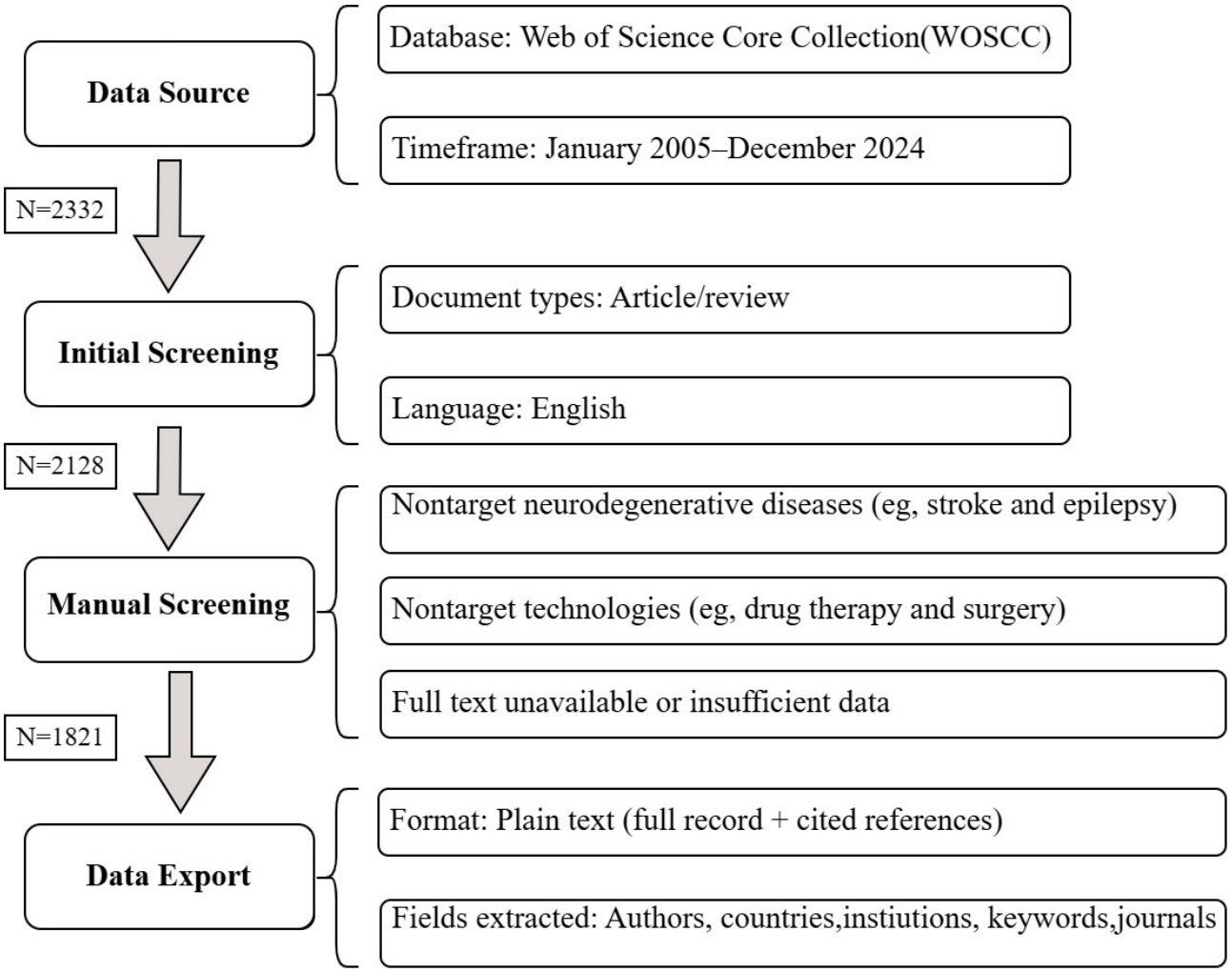
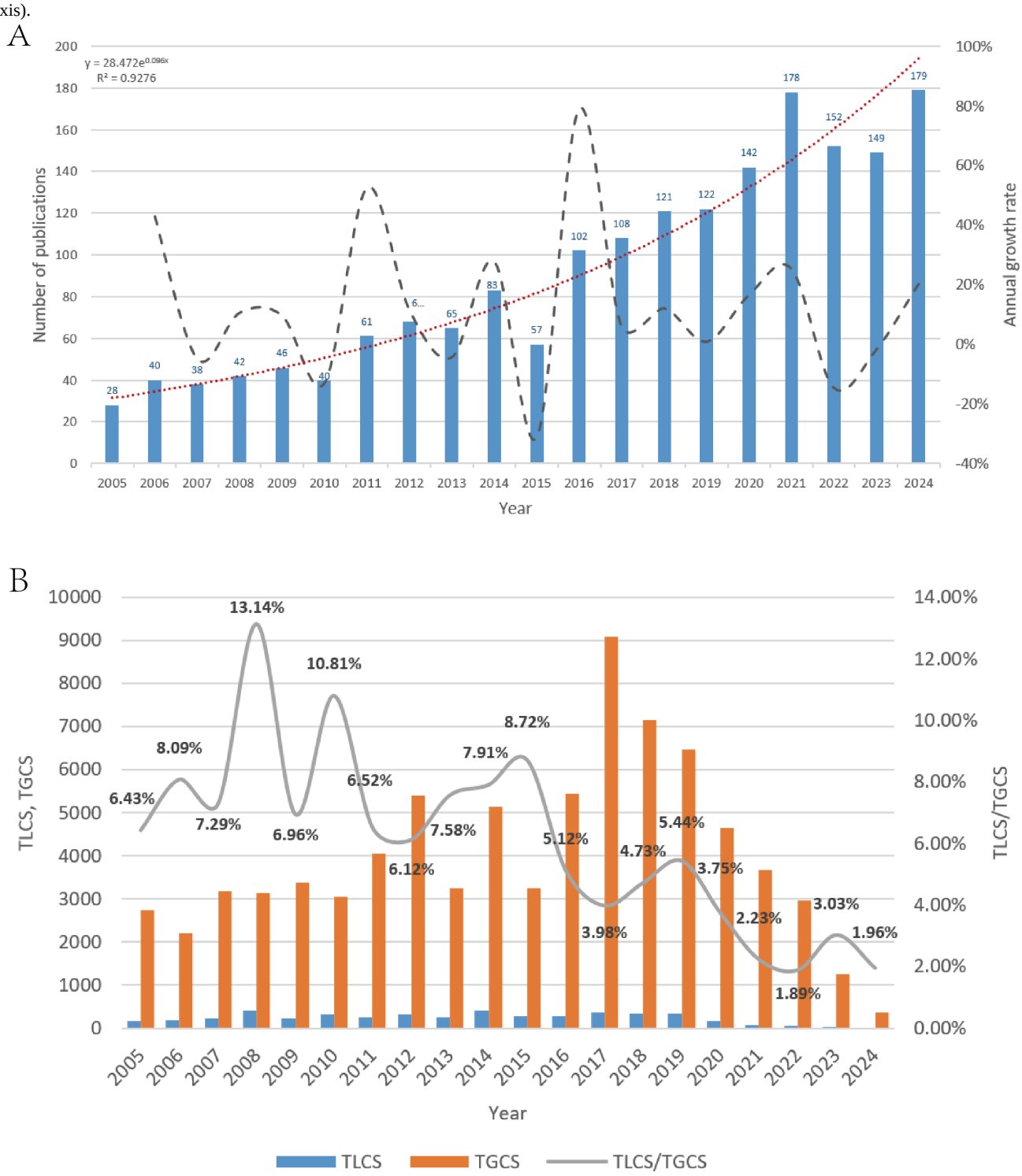



Figure 2A summarizes the annual publication output, fitted trend, and growth rate from 2005 to 2024. Output increased from 28 publications in 2005 to 179 in 2024, with a stepwise rise over time. This trajectory is separated into 2 phases: an exploratory period (2005 - 2015) with a mean of 57 publications per year, followed by an acceleration period (2016 - 2024) with a mean of 125 publications per year. Figure 2B shows the annual TLCS, annual TGCS, and TLCS/TGCS ratio over the same time window. The annual TLCS peaked in 2012 (TLCS=413), coinciding with the highest TLCS/TGCS ratio (13.14%). In contrast, the annual TGCS reached its maximum value in 2017 (TGCS=9092), reflecting a period in which publications from this field achieved particularly broad citation uptake.



XSL•FO
RenderX

**Figure 2.** Annual publication trend. (A) Annual publication output, trend line, and growth rate from 2005 to 2024. Line chart showing the number of publications per year (bars, left y-axis) and the calculated annual growth rate percentage (line, right y-axis). The dashed trend line is fitted to the publication count. (B) Temporal distribution of the total local citation score (TLCS), total global citation score (TGCS), and TLCS/TGCS ratio. Blue bars represent the annual TLCS (left y-axis), orange bars represent the annual TGCS (left y-axis), and the line represents the TLCS/TGCS ratio (right y-axis).




## Country Analysis

A total of 90 countries published work on gene editing and/or stem cell therapies for NDs. Table 1 summarizes the 10 most productive countries, including the publication count, share, TLCS, TGCS, and average number of citations per article. The United States ranked first by output (n=780) and by citation impact (TLCS=2784; TGCS=40,009), accounting for more than 40% of all publications in the dataset. Germany had the highest average number of citations per article (97.90). China and India ranked among the top 5 in publication volume but had the lowest average number of citations per article among leading countries, indicating high throughput but comparatively limited per-paper impact.



XSL•FO
RenderX

**Table .** Top 10 countries with the most publications.

| Rank | Country | Documents (N=1821), n (%) | TLCS[a] | TGCS[b] | Average article citation |
|---|---|---|---|---|---|
| 1 | United States | 780 (42.8) | 2784 | 40,009 | 51.29 |
| 2 | China | 282 (15.5) | 384 | 8180 | 29.01 |
| 3 | Italy | 146 (8.0) | 448 | 6807 | 46.62 |
| 4 | United Kingdom | 138 (7.6) | 262 | 9980 | 72.32 |
| 5 | India | 110 (6.0) | 70 | 2159 | 19.63 |
| 6 | France | 91 (5.0) | 214 | 4684 | 51.47 |
| 7 | Japan | 88 (4.8) | 146 | 2462 | 27.98 |
| 8 | Germany | 80 (4.4) | 102 | 7832 | 97.90 |
| 9 | South Korea | 79 (4.3) | 257 | 2890 | 36.58 |
| 10 | Canada | 77 (4.2) | 159 | 6690 | 86.88 |

[a]TLCS: total local citation score.

[b]TGCS: total global citation score.

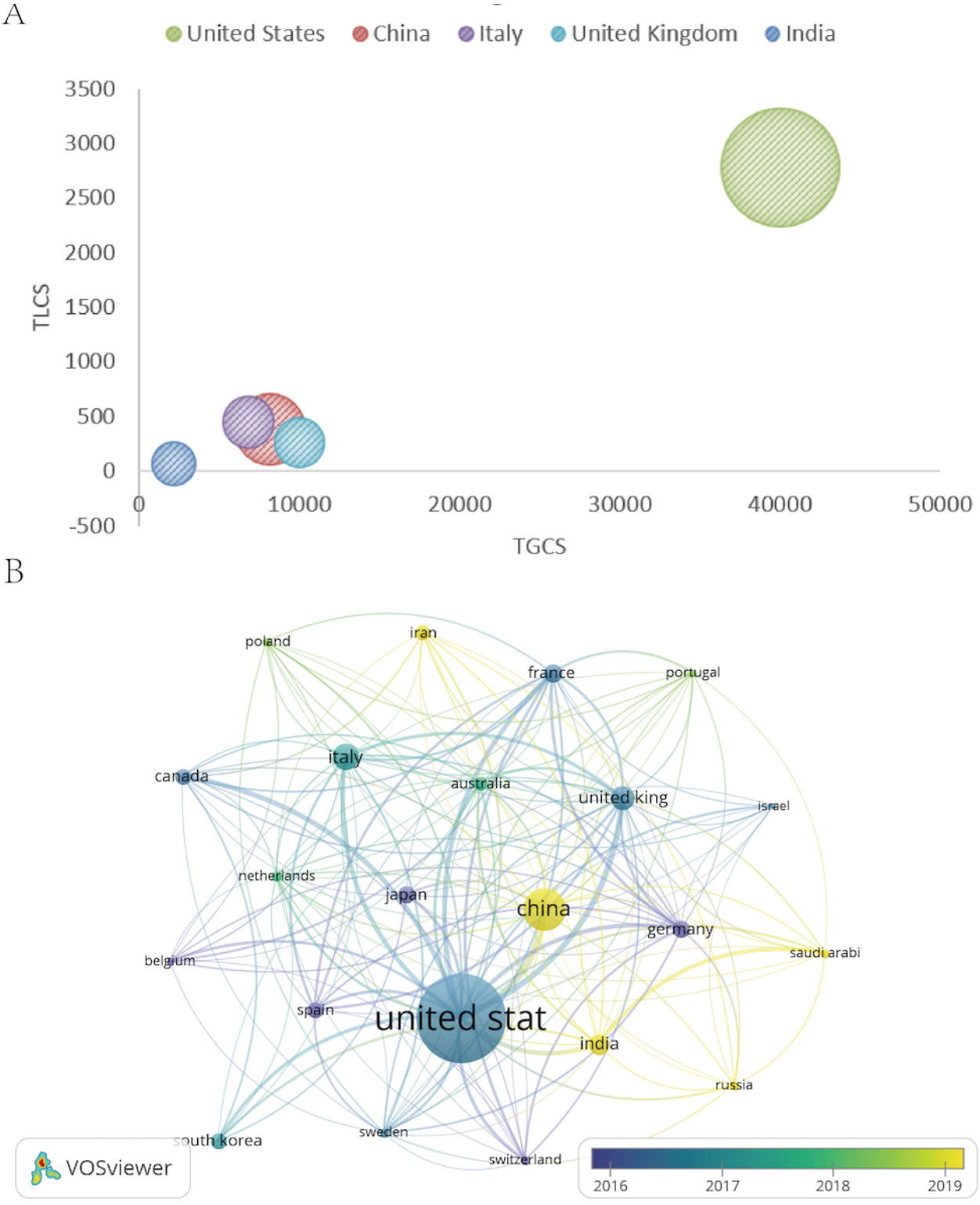
Figure 3A visualizes the top 5 countries using a bubble plot in which the TGCS (cross-field reach) is shown on the x-axis and the TLCS (within-field influence) is shown on the y-axis; bubble size reflects publication volume. The United States clustered in the upper-right quadrant, consistent with strong performance across productivity and both citation dimensions. China ranked second by output but remained well below the United States in the TLCS and TGCS, suggesting that publication volume has outpaced citation influence. Italy and the United Kingdom produced similar numbers of papers but showed distinct citation profiles: the United Kingdom had a higher TGCS (9980 vs 6807), whereas Italy had a higher TLCS (448 vs 262), consistent with differences in how research from each country disseminates within versus beyond the core field. India showed a lower TLCS and TGCS overall, which may reflect differences in infrastructure, funding, and international visibility.

Country collaboration patterns are shown in the overlay visualization (Figure 3B). Nodes denote countries, edges indicate collaboration strength, node size represents publication volume, and node color reflects average publication year. The network reveals broad international connectivity, with the United States serving as a central hub. Several late-emerging contributors, including China, India, Iran, Saudi Arabia, and Russia, showed more recent activity, suggesting a gradual diversification of global participation in this area.



XSL•FO
RenderX

**Figure 3.** Country analysis. (A) Bubble chart of the top 5 countries by publication volume. The x-axis represents the total global citation score (TGCS) as a measure of cross-disciplinary reach, while the y-axis represents the total local citation score (TLCS), indicating field-specific recognition. The bubble size corresponds to the total publication count. (B) Country overlay visualization diagram. Nodes represent countries, with size proportional to the publication volume. Links indicate collaboration between countries. Node color denotes the average publication year.




## Analysis of Research Institutions

A total of 2515 institutions contributed to research on gene editing and/or stem cell therapies for NDs. The 5 most productive institutions are listed in Table 2; all these are based in the United States, consistent with the country-level dominance observed above. Publication counts across these top institutions were closely clustered (32 - 41 papers), suggesting broad, parallel investment rather than concentration within a single "monopoly" center. Institution-level impact differed substantially. As shown in Figure 4A, the University of California, San Francisco, showed the highest average citations per paper (153.15), nearly threefold higher than Johns Hopkins University (55.73). This gap indicates that the University of



XSL•FO
RenderX

California, San Francisco, not only had a high-volume output but also contributed disproportionately influential work within the field. Collaboration patterns are shown in Figure 4B. The national network highlights several strong institutional pairings, including Emory University-University of Michigan, Harvard University-University of California, San Francisco, and Massachusetts General Hospital–associated nodes. Together, these connections suggest an ecosystem in which basic research, technology development, and clinical programs are interlinked, supporting a "bench-to-clinic" pipeline. The overlay map (Figure 4C) further indicates a shift toward broader participation over time, with institutions, such as the University of Massachusetts, Harvard Medical School, the University of Pennsylvania, University College London, and the Chinese Academy of Sciences, appearing as increasingly active contributors in more recent years.

**Table .** Top 5 institutions with the most publications.

| Rank | Institution | Country | Documents, n | Average article citation | TLCS[a] | TGCS[b] |
|---|---|---|---|---|---|---|
| 1 | Emory University | United States | 41 | 49.34 | 310 | 2023 |
| 2 | University of Massachusetts | United States | 41 | 34.51 | 159 | 1415 |
| 3 | Johns Hopkins University | United States | 33 | 55.73 | 129 | 1839 |
| 4 | University of California, San Francisco | United States | 33 | 153.15 | 264 | 5054 |
| 5 | Harvard Medical School | United States | 32 | 32.31 | 69 | 1034 |

[a]TLCS: total local citation score.

[b]TGCS: total global citation score.



XSL•FO
RenderX

**Figure 4.** Research institution analysis. (A) The leading 5 institutions by publication volume and average citation impact. (B) Institutional collaboration network visualization diagram. Nodes represent institutions, with size proportional to the publication volume. Edges indicate co-authorship links, with thickness weighted by collaboration frequency. Only institutions with ≥20 publications are shown. (C) Institutional overlay visualization diagram. Overlay visualization map showing institutions colored by their average publication year, indicating shifts in research activity over time.



XSL•FO
RenderX

### Analysis of Research Authors

Across the 1821 publications, 9978 authors contributed to this research area. The 5 most productive authors are listed in Table 3. Boulis NM ranked first by publication count (24 papers) and contributed a highly cited clinical study in the *Annals of Neurology* (TGCS=157), "Intraspinal neural stem cell transplantation in ALS: phase 1 trial outcomes," which established feasibility and safety for intraspinal neural stem cell delivery in ALS and helped shape subsequent translational efforts in the field [21]. Bankiewicz KS ranked second and showed the highest average number of citations per paper (124.07), consistent with sustained influence from work on aromatic L-amino acid decarboxylase (AADC) gene therapy for PD [22]. Feldman EL showed the highest H-index (104), reflecting broad and durable citation impact across multiple ND-related themes. Author collaboration is visualized in Figure 5A and B. Node sizes reflect publication volume, and the relatively small differences among the top authors point to a distributed authorship landscape rather than dominance by a single group. The overlay map further suggests increased activity around 2017, marked by the appearance of multiple highly active contributors (eg, Mazzini Leti, Davidson Bev, and Hetz Claudio), consistent with a period of accelerated field growth and diversification of research directions.

**Table .** Top 5 authors with the most publications.

| Rank | Author | Documents, n | TLCS[a] | TGCS[b] | Average article citation | H-index | Organization |
|---|---|---|---|---|---|---|---|
| 1 | Boulis NM | 24 | 223 | 1106 | 46.08 | 32 | Emory University |
| 2 | Bankiewicz KS | 14 | 267 | 1737 | 124.07 | 68 | Ohio State University |
| 3 | Feldman EL | 18 | 263 | 1064 | 59.11 | 104 | University of Michigan |
| 4 | Svendsen CN | 16 | 213 | 970 | 60.63 | 76 | Cedars-Sinai Medical Center |
| 5 | Federici T | 12 | 195 | 614 | 51.17 | 54 | University of Milan |

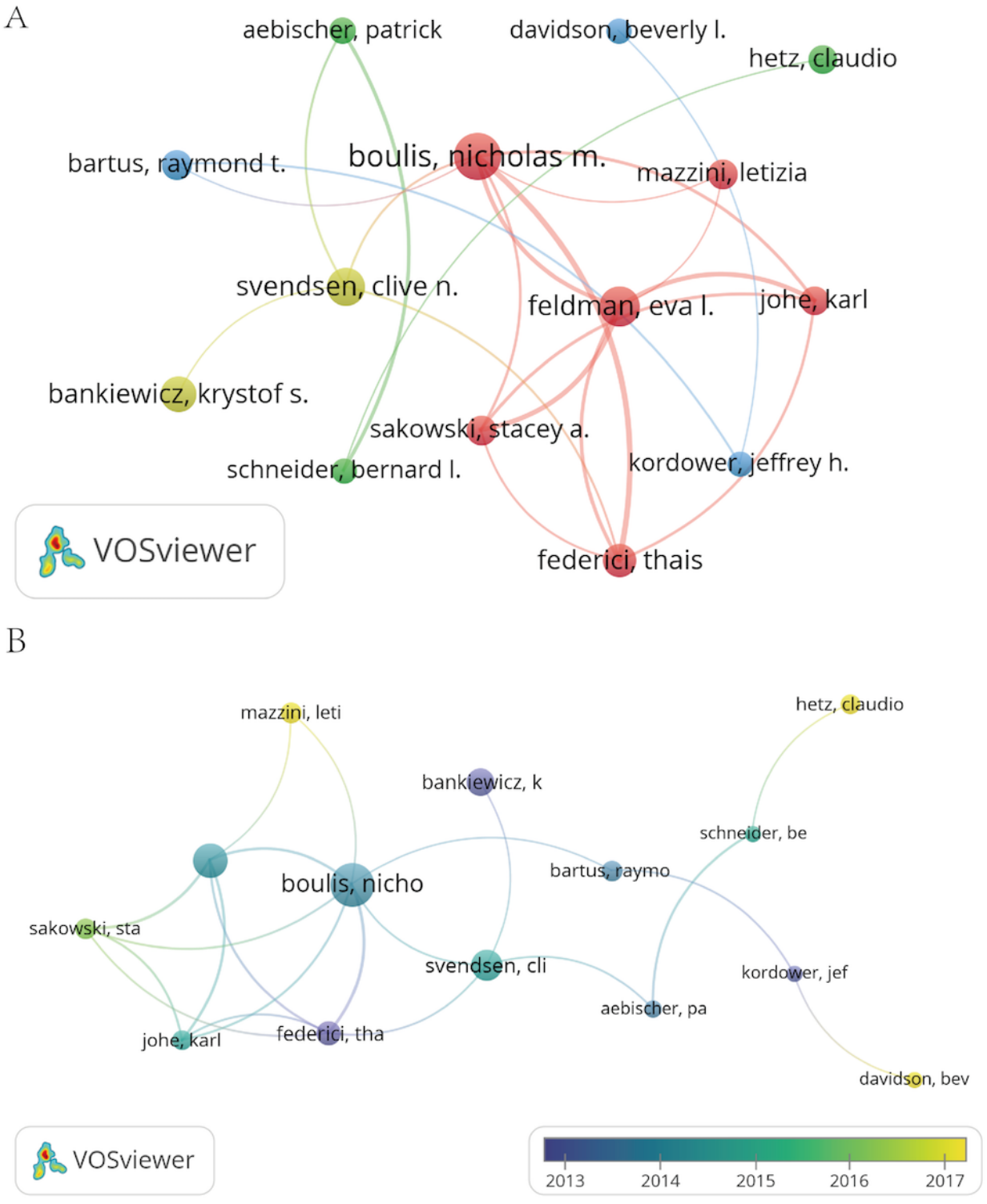
[a]TLCS: total local citation score.

[b]TGCS: total global citation score.



XSL•FO
RenderX

**Figure 5.** Research author analysis. (A) Author collaboration network visualization diagram. Nodes represent authors, sized by their publication count, and edges connect authors who have co-authored publications. Labels are shown for authors with ≥10 publications. (B) Author overlay visualization diagram. Overlay visualization of the author co-authorship network, with node color representing the average publication year of each author (gradient: blue=older, yellow=recent), highlighting authors active in different periods.




## Journal Analysis

The 1821 publications were distributed across 649 journals. The 5 most productive journals are summarized in Table 4. Collectively, these data illustrate a stratified dissemination landscape in which publication volume and citation influence do not always align. The *International Journal of Molecular Sciences* published the highest number of papers (n=52), but showed a modest impact factor (4.9) relative to other core journals, alongside 1586 total citations and an H-index of 114. This pattern suggests high throughput with more variable per-paper influence. In contrast, *Molecular Therapy* combined substantial output (45 papers) with the strongest citation profile among the leading journals (4300 total citations; impact factor



XSL•FO
RenderX

of 12.1; H-index of 158), supporting its position as a central venue for high-impact work in this domain. *Neurobiology of Disease* ranked third (30 papers; impact factor of 5.1; 1327 citations; H-index of 151), consistent with sustained influence from mechanistic and disease-model studies. *PLOS One* ranked fourth (29 papers; impact factor of 2.9; 996 citations; H-index of 268), reflecting broad topical coverage and extensive long-term citation accumulation. *Gene Therapy* ranked fifth (27 papers; impact factor of 4.6; 1179 citations; H-index of 153), maintaining a stable niche within gene therapy research.

**Table .** Top 5 journals by publication volume.

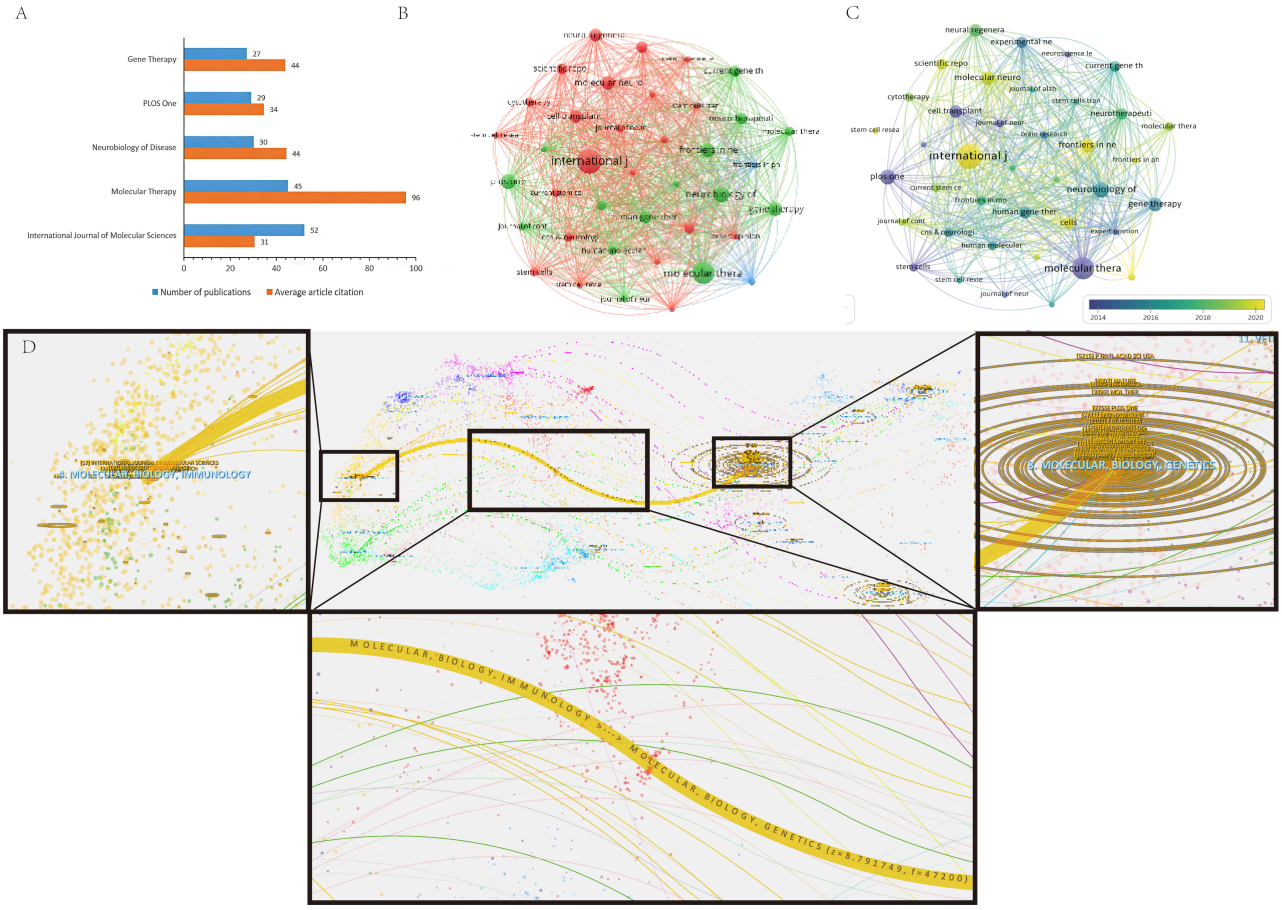
| Rank | Journal | Total publications, n | Total citations, n | Impact factor | H-index |
|---|---|---|---|---|---|
| 1 | *International Journal of Molecular Sciences* | 52 | 1586 | 4.9 | 114 |
| 2 | *Molecular Therapy* | 45 | 4300 | 12.1 | 158 |
| 3 | *Neurobiology of Disease* | 30 | 1327 | 5.1 | 151 |
| 4 | *PLOS One* | 29 | 996 | 2.9 | 268 |
| 5 | *Gene Therapy* | 27 | 1179 | 4.6 | 153 |

Figure 6A compares publication volume and mean citations per paper for these journals and again highlights *Molecular Therapy* as a top contributor by both output and influence (45 papers; 95.56 citations per paper). Its most cited paper (226 citations), "Direct muscle delivery of glial cell line–derived neurotrophic factor (GDNF) via human mesenchymal stem cells (MSCs) improves motor neuron survival and function in a rat model of familial ALS," reported the therapeutic benefit of GDNF-expressing human MSCs in an ALS rat model and serves as a representative translational milestone in this space [23]. Figure 6B shows dense interjournal citation connectivity, consistent with rapid knowledge turnover and active cross-referencing across core venues. The dual-map overlay (Figure 6C) indicates directional citation flows from citing journals focused on molecular biology and immunology toward citing journals focused on molecular biology and genetics, supporting a convergence of delivery/immune considerations with genetic and mechanistic foundations.



XSL•FO
RenderX

**Figure 6.** Journal analysis. (A) Top 5 journals by publication volume and field-weighted citation impact. Journal names are listed on the vertical axis. Blue bars represent the total number of publications per journal. Orange bars represent the average citations per article for that journal within the analyzed dataset. (B) Journal collaboration network visualization diagram. (C) Journal overlay visualization diagram. (D) Dual-map overlay of journals. Left-side clusters (citing journals) are concentrated in molecular biology/immunology and health/medicine, while right-side clusters (cited journals) are focused on molecular biology/genetics and clinical medicine. Colored curves indicate citation links, with curve thickness proportional to citation frequency (f). Key metrics: z=8.79, f=47,200.




## Literature Co-Citation Analysis

To identify foundational studies shaping the field, we examined highly influential documents by the TLCS. Table 5 lists the 5 papers with the highest TLCS, which collectively span gene therapy approaches (including adeno-associated virus [AAV]–mediated glutamate decarboxylase [GAD]/AADC delivery) and stem cell–based strategies (including MSC transplantation). These studies also represent a range of administration routes (eg, intrathecal, intravenous, and intracerebral delivery) and translational stages from animal models to phase I clinical testing.



XSL•FO
RenderX

**Table .** Top 5 publications with the highest local total citation score.

| Rank | Title | Journal | First author | Year | TLCS[a] |
|---|---|---|---|---|---|
| 1 | A phase 1 clinical trial of nerve growth factor gene therapy for Alzheimer disease [24] | *Nature Medicine* | Tuszynski MH | 2005 | 102 |
| 2 | Safety and tolerability of gene therapy with an adeno-associated virus (AAV) borne GAD gene for Parkinson's disease: an open label, phase I trial [25] | *Lancet* | Kaplitt MG | 2007 | 95 |
| 3 | Mesenchymal stem cell transplantation in amyotrophic lateral sclerosis: a phase I clinical trial [26] | *Experimental Neurology* | Mazzini L | 2010 | 86 |
| 4 | Safety and immunological effects of mesenchymal stem cell transplantation in patients with multiple sclerosis and amyotrophic lateral sclerosis [27] | *Archives of Neurology* | Karussis D | 2010 | 71 |
| 5 | Safety and tolerability of putaminal AADC gene therapy for Parkinson disease [22] | *Neurology* | Christine CW | 2009 | 65 |

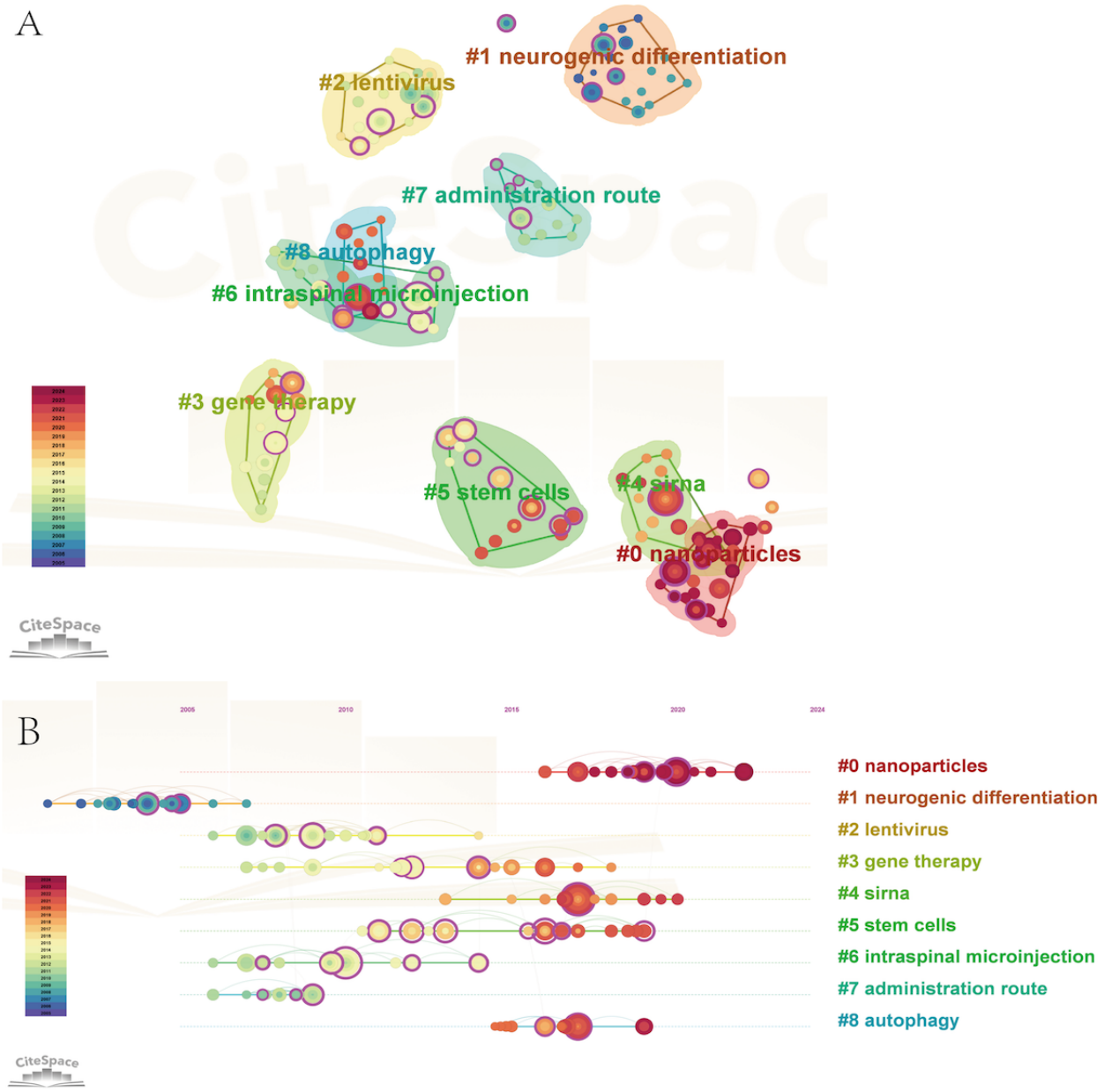
[a]TLCS: total local citation score.

CiteSpace co-citation clustering revealed a strong network structure and cluster coherence (Q=0.8485; S=0.9513; Figure 7A). Nine major clusters were identified, including "nanoparticles," "neurogenic differentiation," "lentivirus," "gene therapy," "siRNA," "stem cells," "intraspinal microinjection," "administration route," and "autophagy," collectively outlining the core intellectual pillars of the field. The timeline view (Figure 7B) further suggests that "nanoparticles," "siRNA," and "autophagy" have become more prominent in recent years, consistent with increasing emphasis on delivery engineering and mechanistic modulation.



XSL•FO
RenderX

**Figure 7.** Co-citation analysis. (A) Clustering diagram of co-cited documents. Network of co-cited references clustered using CiteSpace. Nodes represent references, and links represent co-citation strength. Colors denote distinct thematic clusters (see cluster labels #0-#8). Clustering quality metrics: modularity Q=0.8485; mean silhouette S=0.9513. (B) Time evolution diagram of clustering of co-cited documents. Timeline visualization of the co-citation clusters shown in part A. The x-axis represents the publication year (2005-2024). Each horizontal band corresponds to a cluster, showing the timespan during which its constituent references were cited. Clusters are labeled #0-#8.



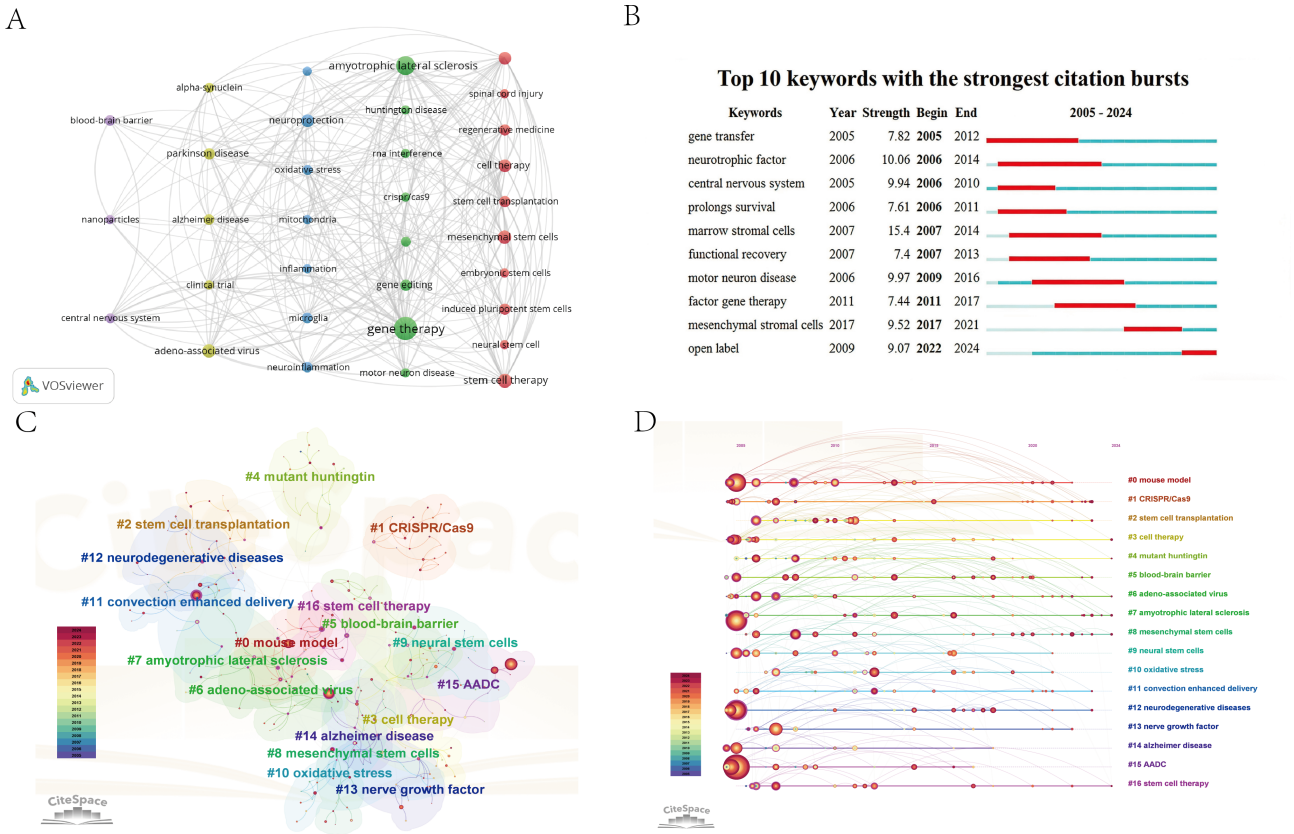

## Keyword Analysis

Keywords provide a compact representation of study focus and are well-suited for tracking thematic evolution. Figure 8A shows 31 keywords occurring more than 20 times, highlighting recurring attention to AAV, autophagy, blood-brain barrier (BBB), CRISPR/Cas9, iPSCs, and nanoparticles. Keyword co-occurrence analysis yielded high clustering quality (Q=0.7813; S=0.9197), indicating that the identified keyword groups are both well separated and internally consistent.



XSL•FO
RenderX

**Figure 8.** Keyword analysis. (A) Keyword co-occurrence network diagram. Node size corresponds to keyword frequency. Keywords are grouped by color based on co-occurrence patterns. (B) Keyword burst diagram. Burst detection analysis performed in CiteSpace (γ=0.5). Each horizontal bar represents a keyword, with red segments indicating the period of significant frequency increase ("burst"). (C) Keyword clustering diagram. Network of co-occurring keywords clustered using CiteSpace. Colors denote 17 distinct thematic clusters (labeled #0-#16). Clustering quality metrics: modularity Q=0.7813; mean silhouette S=0.9197. (D) Keyword clustering time evolution diagram. Timeline view of the keyword clusters from part C, showing the activity period of each cluster from 2005 to 2024. The x-axis represents time, and each horizontal line corresponds to a cluster (#0-#16), visualizing the emergence and persistence of research themes.




Burst detection (Figure 8B) identified "marrow stromal cells" as the strongest burst term (strength=15.4), while "neurotrophic factor" showed the longest burst duration (2006 - 2014), consistent with an early wave of neuroprotection-focused strategies. Notably, "open label" emerged as a recent burst term, aligning with increasing attention to early-stage clinical translation.

Keyword clustering further resolved 17 clusters (Figure 8C), spanning disease entities and mechanisms (eg, NDs, ALS, AD, mutant huntingtin, and oxidative stress), gene editing and gene therapy platforms (eg, CRISPR/Cas9, AAV, and AADC), and stem cell or cell therapy strategies (eg, stem cell transplantation, MSCs, and neural stem cells). Additional clusters captured experimental models and neuroprotection-related themes (eg, mouse model and nerve growth factor) as well as delivery paradigms (eg, BBB and convection-enhanced delivery [CED]). The temporal map (Figure 8D) indicates sustained activity across multiple directions over the past 2 decades, with concurrent advances in delivery optimization, mechanistic deepening, and clinical translation. Together, these trends point to a field with not only strong foundational accumulation but also persistent bottlenecks that continue to motivate technological innovation and cross-disciplinary integration.

## Discussion

The treatment of NDs is at a critical inflection point, with gene editing and stem cell therapy increasingly converging as translational strategies. To our knowledge, this study is the first bibliometric analysis to systematically map the global research landscape and emerging trends in this field. Over the past 2 decades, annual publication output increased in a stepwise fashion and entered a clear acceleration phase after 2016. This shift likely reflects several enabling advances in gene and cell therapy. For example, the rAAV2-retro variant, which has been generated through AAV capsid engineering, can efficiently access projection neurons via retrograde transport, improving gene delivery performance [28]. In parallel, iPSCs became essential for disease modeling, and MSCs engineered to overexpress brain-derived neurotrophic factor (BDNF) improved outcomes in HD mouse models [29]. Together, these developments likely contributed to sustained growth from 2016 onward. High-impact outputs in this area also show recognizable milestone patterns. The local citation peak in 2008 (TLCS=413; TLCS/TGCS=13.14%) may relate to studies reporting that MSC transplantation can prolong survival, reduce neuroinflammation, and improve motor neuron function in ALS mouse models [30,31]. In contrast, the global citation peak in 2017 (TGCS=9092) likely reflects broad uptake of the widely cited review "Parkinson disease" in *Nature Reviews Disease Primers*



XSL•FO
RenderX

(2017), which synthesized gene therapy targets and vectors as well as the trajectory and current status of stem cell–based strategies in PD [32]. These peaks appear to align with 2 key stages of field development: proof-of-concept experimental validation and consolidation of translational pathways.

At the country level, the United States maintains a dominant position in publication output as well as in the TLCS and TGCS, and the top 5 institutions are all based in the United States. The Boston area, anchored by Harvard University and Massachusetts General Hospital, exemplifies a "basic-to-clinical" collaborative ecosystem, where geographic proximity may facilitate high-frequency interaction and accelerate translation. The sustained US advantage likely reflects multiple reinforcing factors. Long-term and stable research and development investment provides resilience for high-risk, long-cycle research and supports continuity in key directions. In addition, policy and institutional infrastructure (eg, federal funding mechanisms, evaluation systems, open science practices, and industry-academia translation pathways) may improve efficiency across the pipeline—from project initiation to dissemination—thereby strengthening both output and impact. Finally, the centrality of US institutions within international collaboration networks increases access to high-impact consortia and may amplify visibility and downstream citation performance.

China and India contribute substantial publication volume, but their per-paper citation impact remains comparatively low (China: 248 articles; 29.01 citations per article; India: 110 articles; 19.63 citations per article). Collaboration structure may partly explain this gap. China's collaboration intensity with Western countries remains far below that of the United States (Figure 3B), and India appears more peripheral in the international network, potentially limiting exchange and dissemination. Journal distribution may also contribute. As shown in Table 4 and Figure 6A, outputs from these countries more often appear in high-volume journals with lower field-specific influence, whereas representation in specialty high-impact venues, such as *Molecular Therapy*, remains limited. In addition, keyword evolution (Figures 7 and 8) suggests that research from these countries may concentrate more heavily on applied and translational clusters, while fewer outputs appear to define foundational, concept-setting advances that often drive long-term citation accumulation. Collectively, these patterns suggest that increasing influence will require deeper integration into high-impact international networks and more sustained investment in originality and frontier research.

At the same time, our data suggest a complementary global innovation landscape. The United States shows broad leadership with strong representation in foundational and early-translational clusters, such as novel AAV engineering and first-in-human studies, consistent with its network position and presence in high-impact journals [33]. In contrast, publication growth from countries, such as China and India, aligns more strongly with applied technology clusters, including nanoparticle delivery systems and stem cell transplantation protocols [34]. European contributors, including Germany and the United Kingdom, show sustained depth in mechanistic clusters, such as oxidative stress and protein aggregation, reflecting enduring strengths in fundamental biomedical science [35]. These patterns argue for collaboration models that pair mechanistic depth, platform engineering, and clinical execution. Institutions seeking higher impact may benefit from long-term partnerships with leading laboratories through co-designed projects and joint publication strategies, alongside deliberate targeting of field-leading specialty journals and sustained investment in frontier directions.

Our co-citation and keyword analyses highlight recurring focal areas, including "adeno-associated virus," "nanoparticles," "blood-brain barrier," and "marrow stromal cells." AAV vectors remain central tools due to relatively low immunogenicity, broad tropism, and stable long-term transgene expression [36]. We also identified clusters, such as "lentivirus" and "siRNA." Lentiviral vectors can enable durable gene expression through genomic integration and provide relatively large cargo capacity for platforms such as CRISPR/Cas9 or neurotrophic factor genes [37,38]. Small interfering RNA (siRNA) can silence pathogenic genes through RNA interference and may also complement gene-editing strategies by modulating target expression or reducing unintended activity [39]. However, clinical translation remains constrained by 2 major bottlenecks [40]. First, BBB penetration is often insufficient in humans, despite robust performance in some animal studies. Second, neutralizing antibodies against AAV can reduce transduction efficiency and may contribute to treatment failure, while immune-evasion strategies remain comparatively underdeveloped. Recent progress in surface-modified and targeted nanoparticles has improved BBB delivery [41]. In our analyses, "nanoparticles" emerged as both a persistent focus and a growing frontier. Nanoparticles (1 - 100 nm) can be engineered (eg, using rabies virus glycoprotein peptides or lactoferrin) to facilitate BBB transport; protect payloads, such as AAV and siRNA; and potentially influence stem cell survival and differentiation [42,43]. Encapsulating AAV within lipid nanoparticles, for example, has been reported to enhance vector stability and in vivo penetration through optimization of lipid composition and structure [44]. These advances underscore the promise of integrating material science with vector biology to improve delivery and the therapeutic index.

Another prominent cluster was "autophagy," a cellular quality-control pathway that clears misfolded proteins and damaged organelles. Enhancing autophagy is an attractive therapeutic direction, but precise bidirectional control remains challenging because excessive activation risks degrading essential proteins [45]. MSCs and MSC-derived extracellular vesicles also remain highly visible in the literature. The burst term "marrow stromal cells" supports sustained attention to MSC-based strategies. Bone marrow–derived MSCs can support neural repair through the secretion of neurotrophic factors and immunomodulatory effects [46]. Human MSCs engineered to overexpress BDNF improved outcomes in HD mouse models [29,47]. Laromestrocel, an allogeneic bone marrow–derived MSC product from young healthy donors, has shown encouraging findings in AD, including reduced rates of whole-brain volume loss and hippocampal atrophy, without amyloid-related imaging abnormalities during treatment [48]. As a cell-free alternative, MSC-derived exosomes are emerging as nanoscale carriers capable of delivering therapeutics to the



XSL•FO
RenderX

brain, supporting Aβ clearance and neuroprotection [49]. MSC-derived exosomes have also been reported to localize to inflamed brain regions and to be taken up by neurons, suggesting potential for targeted delivery [50]. The recent emergence of "open label" indicates that clinical translation is advancing but still concentrated in early validation. Consistent with this, many highly cited clinical studies in this field are phase I open-label trials. While open-label designs are practical for early exploration, definitive efficacy requires subsequent confirmation in larger randomized, double-blind controlled studies [24-27]. Beyond vector-related barriers, translation also faces platform-wide challenges: nanoparticle safety and long-term stability in humans remain incompletely defined; MSC survival and integration after transplantation are often limited; gene editing carries the risk of off-target effects; and stem cell–based approaches raise concerns regarding tumorigenicity and immune reactions. Finally, many trials remain limited by a small sample size and a short follow-up, underscoring the need for larger cohorts and longer-term surveillance to establish durability and rare safety signals.

Our keyword clustering further resolved thematic dimensions beyond tools and cells. These included disease entities and mechanisms (eg, oxidative stress, mutant huntingtin, AD, and ALS), gene-editing and gene-therapy platforms (CRISPR/Cas9, AAV, and AADC), stem cell and cell therapy strategies (MSCs, neural stem cells, and transplantation), delivery paradigms (CED and BBB), and experimental models and neuroprotection-related themes (mouse models and nerve growth factor). Oxidative stress reflects cellular injury driven by excessive reactive oxygen and nitrogen species, and interventions that neutralize reactive oxygen species remain under active exploration, although deeper mechanistic work is still needed to guide therapeutic development. CRISPR/Cas9 enables precise genome manipulation [51], and AADC gene therapy has shown translational potential in PD, although dose optimization and immunogenicity remain important challenges. Stem cell therapy offers multitarget potential but continues to face constraints in standardized manufacturing, delivery efficiency, and long-term safety [52-56]. Delivery remains a central determinant of success: CED can improve distribution within the central nervous system and offers a strategy to address BBB-associated constraints [57,58]. Animal models remain essential, but differences between rodents and humans, particularly BBB properties and lifespan, can contribute to translational inconsistency. Clinical studies of neurotrophic factors, such as nerve growth factor, have generally shown safety but limited efficacy in late-stage disease, highlighting the complexity of these interventions [59]. Overall, progress will depend on improving delivery and efficiency, reducing immune barriers, and strengthening safety assessments to support reliable clinical benefit.

The integration of iPSC technology with CRISPR/Cas9 editing remains particularly promising for personalized approaches. Patient-derived iPSCs can mitigate immune rejection concerns while enabling gene correction or targeted modification. In HD, CRISPR/Cas9 correction of CAG-repeat expansions in *HTT* in patient-derived iPSCs yielded differentiated neurons with improved functional phenotypes [60]. Additionally, iPSCs can differentiate into diverse neural cell types, and when combined with CRISPR/Cas9 technology, they enable the generation of specialized cells tailored for multicell-type ND therapies. In PD treatment, researchers differentiated iPSCs into dopaminergic neurons, applied CRISPR-mediated gene correction, and observed symptom improvement in PD models following transplantation [61]. In AD, neurons or astrocytes can be obtained through differentiation of patient-derived iPSCs, and CRISPR/Cas9 can be used for precise knock-in/out or allele replacement of key risk genes, such as *APP*, *PSEN1/2*, and *APOE*, thereby reversing most phenotypes [62]. In ALS, pathogenic variants around *SOD1*, *C9ORF72*, *TARDBP*, and *FUS* can be investigated in motor neurons differentiated from iPSCs and their co-culture systems, observing key phenotypes such as RNA metabolism abnormalities, stress granules, axonal transport deficits, and cell nonautonomous damage, as well as using gene correction to validate the causal relationship between phenotypes and mutations [63]. Across multiple ND models, combined iPSC-CRISPR strategies have shown therapeutic benefit with acceptable tolerability and low immunogenicity in animal studies [64].

In the process of transitioning technologies from the laboratory to clinical and industrial applications, ethical considerations and regulatory frameworks often pose more immediate practical bottlenecks, impacting whether new technologies can gain approval, establish trust, and achieve large-scale implementation, in addition to scientific feasibility and engineering maturity. Ethically, it is essential to regulate informed consent, privacy, and data security while ensuring fairness in the risk-benefit balance. In the context of iPSCs and CRISPR, particular attention should be given to evaluating off-target or unintended effects, long-term safety, follow-up requirements, and the protection of vulnerable populations. On the regulatory front, this involves product attributes, quality systems, and the requirements for clinical evidence and risk management, including endpoint settings and long-term safety evaluations. These factors significantly influence research design, data availability, timelines, and costs, ultimately determining the efficiency of translation.

This study has several limitations. First, literature retrieval was restricted to the WoSCC, which may omit journals or regional contributions not indexed in this database. Second, although search terms were refined, omissions cannot be fully excluded. Third, we did not incorporate additional translational indicators, such as patent landscapes.

This study systematically reveals dynamic development trends and research hotspots in gene editing and stem cell therapy for NDs through bibliometric analysis. Global annual publications show stepwise growth, with countries like the United States and Germany making outstanding contributions. However, clinical translation faces challenges, such as low BBB penetration efficiency and immune responses, though emerging approaches like nanoparticle-rabies virus glycoprotein peptide modification and MSC exosomes demonstrate breakthrough potential. Additionally, combined applications of iPSCs and CRISPR/Cas9 gene editing show significant therapeutic promise. The "open label" keyword burst indicates that most studies remain in early validation phases, characterized by small clinical trial samples,



XSL•FO
RenderX

short observation periods, and weak evidence levels, necessitating urgent validation of long-term safety and efficacy. Future efforts should prioritize optimizing delivery systems, validating clinical translation, and fostering interdisciplinary research (eg, artificial intelligence–driven AAV capsid design) through dedicated ND funds. Strengthening "basic-clinical-industrial" collaborative networks, enhancing clinical evidence quality, and exploring CRISPR/Cas9-stem cell integration will provide critical pathways to overcoming therapeutic bottlenecks in NDs.

## Acknowledgments

No generative artificial intelligence tools were used in the preparation, writing, analysis, or editing of this manuscript. All work was conducted by the human authors.

## Funding

This study was financially supported by the Zhangjiajie Science and Technology Development Project (number: 202305) and the Wu Jieping Medical Foundation Clinical Research Special Fund (number: 320.6750.2024-13-62).

## Data Availability

All data used in this study were derived from publicly accessible databases and literature resources (Web of Science), with data processing performed using publicly available professional tools (CiteSpace, VOSviewer, and HistCite). Additionally, datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

## Authors' Contributions

JK (Phone: 86 13974499671; Email: Kejun1979@163.com) and GP (Phone: 86 18874042983 Email: 309932540@qq.com) are co-corresponding authors.

## Conflicts of Interest

None declared.

## References

1. Wilson DM III, Cookson MR, Van Den Bosch L, Zetterberg H, Holtzman DM, Dewachter I. Hallmarks of neurodegenerative diseases. Cell 2023 Feb 16;186(4):693-714. [doi: 10.1016/j.cell.2022.12.032] [Medline: 36803602]
2. Erkkinen MG, Kim MO, Geschwind MD. Clinical neurology and epidemiology of the major neurodegenerative diseases. Cold Spring Harb Perspect Biol 2018 Apr 2;10(4):a033118. [doi: 10.1101/cshperspect.a033118] [Medline: 28716886]
3. Spillantini MG, Goedert M. Tau protein pathology in neurodegenerative diseases. Trends Neurosci 1998 Oct;21(10):428-433. [doi: 10.1016/s0166-2236(98)01337-x] [Medline: 9786340]
4. Park H, Kam TI, Dawson VL, Dawson TM. α-Synuclein pathology as a target in neurodegenerative diseases. Nat Rev Neurol 2025 Jan;21(1):32-47. [doi: 10.1038/s41582-024-01043-w] [Medline: 39609631]
5. Cui X, Guo YE, Fang JH, et al. Donepezil, a drug for Alzheimer's disease, promotes oligodendrocyte generation and remyelination. Acta Pharmacol Sin 2019 Nov;40(11):1386-1393. [doi: 10.1038/s41401-018-0206-4] [Medline: 30918344]
6. Olanow CW. Levodopa is the best symptomatic therapy for PD: nothing more, nothing less. Mov Disord 2019 Jun;34(6):812-815. [doi: 10.1002/mds.27690] [Medline: 30990922]
7. Carter JE, Schuchman EH. Gene therapy for neurodegenerative diseases: fact or fiction? Br J Psychiatry 2001 May;178:392-394. [doi: 10.1192/bjp.178.5.392] [Medline: 11331549]
8. Temple S. Advancing cell therapy for neurodegenerative diseases. Cell Stem Cell 2023 May 4;30(5):512-529. [doi: 10.1016/j.stem.2023.03.017] [Medline: 37084729]
9. Song B, Cha Y, Ko S, et al. Human autologous iPSC-derived dopaminergic progenitors restore motor function in Parkinson's disease models. J Clin Invest 2020 Feb 3;130(2):904-920. [doi: 10.1172/JCI130767] [Medline: 31714896]
10. Ye T, Duan Y, Tsang HWS, et al. Efficient manipulation of gene dosage in human iPSCs using CRISPR/Cas9 nickases. Commun Biol 2021 Feb 12;4(1):195. [doi: 10.1038/s42003-021-01722-0] [Medline: 33580208]
11. Yang G, Huang X. Methods and applications of CRISPR/Cas system for genome editing in stem cells. Cell Regen 2019 Dec;8(2):33-41. [doi: 10.1016/j.cr.2019.08.001] [Medline: 31666940]
12. Nakagawa S, Samarasinghe G, Haddaway NR, et al. Research weaving: visualizing the future of research synthesis. Trends Ecol Evol 2019 Mar;34(3):224-238. [doi: 10.1016/j.tree.2018.11.007] [Medline: 30580972]
13. Perrier L, Lightfoot D, Kealey MR, Straus SE, Tricco AC. Knowledge synthesis research: a bibliometric analysis. J Clin Epidemiol 2016 May;73:50-57. [doi: 10.1016/j.jclinepi.2015.02.019] [Medline: 26912123]
14. Huang J, Deng K, Liu Y, Xia M, Lei M, Wu M. Global research trends on innate lymphoid cells in the brain, gut and lung field: a bibliometric and visualized analysis. Front Immunol 2024;15:1336666. [doi: 10.3389/fimmu.2024.1336666] [Medline: 38384457]



XSL•FO
RenderX

15. Liu N, Yu W, Sun M, et al. Research trends and hotspots of ferroptosis in neurodegenerative diseases from 2013 to 2023: a bibliometrics study. Heliyon 2024 Apr 30;10(8):e29418. [doi: 10.1016/j.heliyon.2024.e29418] [Medline: 38638970]
16. Abumalloh RA, Nilashi M, Samad S, et al. Parkinson's disease diagnosis using deep learning: a bibliometric analysis and literature review. Ageing Res Rev 2024 Apr;96:102285. [doi: 10.1016/j.arr.2024.102285] [Medline: 38554785]
17. Liu Y, Huang J, Li E, et al. Analysis of research trends and hot spots on COPD biomarkers from the perspective of bibliometrics. Respir Med 2025;240:108030. [doi: 10.1016/j.rmed.2025.108030] [Medline: 40058665]
18. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010 Aug;84(2):523-538. [doi: 10.1007/s11192-009-0146-3] [Medline: 20585380]
19. Bukar UA, Sayeed MS, Razak SFA, Yogarayan S, Amodu OA, Mahmood RAR. A method for analyzing text using VOSviewer. MethodsX 2023 Dec;11:102339. [doi: 10.1016/j.mex.2023.102339] [Medline: 37693657]
20. Chen C. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. J Am Soc Inf Sci 2006 Feb;57(3):359-377. [doi: 10.1002/asi.20317]
21. Feldman EL, Boulis NM, Hur J, et al. Intraspinal neural stem cell transplantation in amyotrophic lateral sclerosis: phase 1 trial outcomes. Ann Neurol 2014 Mar;75(3):363-373. [doi: 10.1002/ana.24113] [Medline: 24510776]
22. Christine CW, Starr PA, Larson PS, et al. Safety and tolerability of putaminal AADC gene therapy for Parkinson disease. Neurology (ECronicon) 2009 Nov 17;73(20):1662-1669. [doi: 10.1212/WNL.0b013e3181c29356] [Medline: 19828868]
23. Suzuki M, McHugh J, Tork C, et al. Direct muscle delivery of GDNF with human mesenchymal stem cells improves motor neuron survival and function in a rat model of familial ALS. Mol Ther 2008 Dec;16(12):2002-2010. [doi: 10.1038/mt.2008.197] [Medline: 18797452]
24. Tuszynski MH, Thal L, Pay M, et al. A phase 1 clinical trial of nerve growth factor gene therapy for Alzheimer disease. Nat Med 2005 May;11(5):551-555. [doi: 10.1038/nm1239] [Medline: 15852017]
25. Kaplitt MG, Feigin A, Tang C, et al. Safety and tolerability of gene therapy with an adeno-associated virus (AAV) borne GAD gene for Parkinson's disease: an open label, phase I trial. Lancet 2007 Jun 23;369(9579):2097-2105. [doi: 10.1016/S0140-6736(07)60982-9] [Medline: 17586305]
26. Mazzini L, Ferrero I, Luparello V, et al. Mesenchymal stem cell transplantation in amyotrophic lateral sclerosis: a phase I clinical trial. Exp Neurol 2010 May;223(1):229-237. [doi: 10.1016/j.expneurol.2009.08.007] [Medline: 19682989]
27. Karussis D, Karageorgiou C, Vaknin-Dembinsky A, et al. Safety and immunological effects of mesenchymal stem cell transplantation in patients with multiple sclerosis and amyotrophic lateral sclerosis. Arch Neurol 2010 Oct;67(10):1187-1194. [doi: 10.1001/archneurol.2010.248] [Medline: 20937945]
28. Tervo DGR, Hwang BY, Viswanathan S, et al. A designer AAV variant permits efficient retrograde access to projection neurons. Neuron 2016 Oct 19;92(2):372-382. [doi: 10.1016/j.neuron.2016.09.021] [Medline: 27720486]
29. Pollock K, Dahlenburg H, Nelson H, et al. Human mesenchymal stem cells genetically engineered to overexpress brain-derived neurotrophic factor improve outcomes in Huntington's disease mouse models. Mol Ther 2016 May;24(5):965-977. [doi: 10.1038/mt.2016.12] [Medline: 26765769]
30. Vercelli A, Mereuta OM, Garbossa D, et al. Human mesenchymal stem cell transplantation extends survival, improves motor performance and decreases neuroinflammation in mouse model of amyotrophic lateral sclerosis. Neurobiol Dis 2008 Sep;31(3):395-405. [doi: 10.1016/j.nbd.2008.05.016] [Medline: 18586098]
31. Riley J, Glass J, Feldman EL, et al. Intraspinal stem cell transplantation in amyotrophic lateral sclerosis: a phase I trial, cervical microinjection, and final surgical safety outcomes. Neurosurgery 2014 Jan;74(1):77-87. [doi: 10.1227/NEU.0000000000000156] [Medline: 24018694]
32. Poewe W, Seppi K, Tanner CM, et al. Parkinson disease. Nat Rev Dis Primers 2017 Mar 23;3:17013. [doi: 10.1038/nrdp.2017.13] [Medline: 28332488]
33. Mingozzi F, High KA. Immune responses to AAV vectors: overcoming barriers to successful gene therapy. Blood 2013 Jul 4;122(1):23-36. [doi: 10.1182/blood-2013-01-306647] [Medline: 23596044]
34. Meng Y, Sui L, Xu T, Zhao H, Yuan Q, Sun L. Research and application prospect of nanomedicine in kidney disease: a bibliometric analysis from 2003 to 2024. Int J Nanomedicine 2025;20:3007-3030. [doi: 10.2147/IJN.S510016] [Medline: 40093546]
35. Abramov AY, Potapova EV, Dremin VV, Dunaev AV. Interaction of oxidative stress and misfolded proteins in the mechanism of neurodegeneration. Life (Basel) 2020 Jun 30;10(7):101. [doi: 10.3390/life10070101] [Medline: 32629809]
36. Hudry E, Vandenberghe LH. Therapeutic AAV gene transfer to the nervous system: a clinical reality. Neuron 2019 Mar 6;101(5):839-862. [doi: 10.1016/j.neuron.2019.02.017] [Medline: 30844402]
37. Wang X, Ma C, Rodríguez Labrada R, et al. Recent advances in lentiviral vectors for gene therapy. Sci China Life Sci 2021 Nov;64(11):1842-1857. [doi: 10.1007/s11427-021-1952-5] [Medline: 34708326]
38. Croci C, Fasano S, Superchi D, et al. Cerebellar neurons and glial cells are transducible by lentiviral vectors without decrease of cerebellar functions. Dev Neurosci 2006;28(3):216-221. [doi: 10.1159/000091919] [Medline: 16679768]
39. Gonzalez-Alegre P. Therapeutic RNA interference for neurodegenerative diseases: from promise to progress. Pharmacol Ther 2007 Apr;114(1):34-55. [doi: 10.1016/j.pharmthera.2007.01.003] [Medline: 17316816]
40. Colella P, Ronzitti G, Mingozzi F. Emerging issues in AAV-mediated in vivo gene therapy. Mol Ther Methods Clin Dev 2018 Mar;8:87-104. [doi: 10.1016/j.omtm.2017.11.007] [Medline: 29326962]




XSL•FO
RenderX

41. Gong B, Zhuang J, Ji W, et al. The long and the short of current nanomedicines for treating Alzheimer's disease. J Transl Int Med 2022 Dec;10(4):294-296. [doi: 10.2478/jtim-2021-0054] [Medline: 36860633]
42. Liu S, Li H, Xi S, Zhang Y, Sun T. Advancing CNS therapeutics: enhancing neurological disorders with nanoparticle-based gene and enzyme replacement therapies. Int J Nanomedicine 2025;20:1443-1490. [doi: 10.2147/IJN.S457393] [Medline: 39925682]
43. Naimi N, Seyedmirzaei H, Hassannejad Z, Soltani Khaboushan A. Advanced nanoparticle strategies for optimizing RNA therapeutic delivery in neurodegenerative disorders. Biomed Pharmacother 2024 Jun;175:116691. [doi: 10.1016/j.biopha.2024.116691] [Medline: 38713941]
44. Michels A, Ho N, Buchholz CJ. Precision medicine: in vivo CAR therapy as a showcase for receptor-targeted vector platforms. Mol Ther 2022 Jul 6;30(7):2401-2415. [doi: 10.1016/j.ymthe.2022.05.018] [Medline: 35598048]
45. García-Huerta P, Troncoso-Escudero P, Wu D, et al. Insulin-like growth factor 2 (IGF2) protects against Huntington's disease through the extracellular disposal of protein aggregates. Acta Neuropathol 2020 Nov;140(5):737-764. [doi: 10.1007/s00401-020-02183-1] [Medline: 32642868]
46. Song CG, Zhang YZ, Wu HN, et al. Stem cells: a promising candidate to treat neurological disorders. Neural Regen Res 2018 Jul;13(7):1294-1304. [doi: 10.4103/1673-5374.235085] [Medline: 30028342]
47. Gao J, Gunasekar S, Xia ZJ, et al. Gene therapy for CNS disorders: modalities, delivery and translational challenges. Nat Rev Neurosci 2024 Aug;25(8):553-572. [doi: 10.1038/s41583-024-00829-7] [Medline: 38898231]
48. Rash BG, Ramdas KN, Agafonova N, et al. Allogeneic mesenchymal stem cell therapy with laromestrocel in mild Alzheimer's disease: a randomized controlled phase 2a trial. Nat Med 2025 Apr;31(4):1257-1266. [doi: 10.1038/s41591-025-03559-0] [Medline: 40065171]
49. Quan J, Liu Q, Li P, et al. Mesenchymal stem cell exosome therapy: current research status in the treatment of neurodegenerative diseases and the possibility of reversing normal brain aging. Stem Cell Res Ther 2025 Feb 21;16(1):76. [doi: 10.1186/s13287-025-04160-5] [Medline: 39985030]
50. Harrell CR, Volarevic A, Djonov V, Volarevic V. Mesenchymal stem cell-derived exosomes as new remedy for the treatment of neurocognitive disorders. Int J Mol Sci 2021 Feb 1;22(3):1433. [doi: 10.3390/ijms22031433] [Medline: 33535376]
51. Yang S, Chang R, Yang H, et al. CRISPR/Cas9-mediated gene editing ameliorates neurotoxicity in mouse model of Huntington's disease. J Clin Invest 2017 Jun 30;127(7):2719-2724. [doi: 10.1172/JCI92087] [Medline: 28628038]
52. Mao Y, Liu P, Wei J, et al. Exosomes derived from umbilical cord mesenchymal stem cell promote hair regrowth in C57BL6 mice through upregulation of the RAS/ERK signaling pathway. J Transl Int Med 2024 Nov;12(5):478-494. [doi: 10.1515/jtim-2024-0012] [Medline: 39513036]
53. Liu Z, Huang Y, Yang Y, et al. Analysis and prediction of research hotspots and trends in heart failure research. J Transl Int Med 2024 Jun;12(3):263-273. [doi: 10.2478/jtim-2023-0117] [Medline: 39081280]
54. Caprnda M, Kubatka P, Gazdikova K, et al. Immunomodulatory effects of stem cells: therapeutic option for neurodegenerative disorders. Biomed Pharmacother 2017 Jul;91:60-69. [doi: 10.1016/j.biopha.2017.04.034] [Medline: 28448871]
55. Han L, Zhou Y, Zhang R, et al. MicroRNA Let-7f-5p promotes bone marrow mesenchymal stem cells survival by targeting caspase-3 in Alzheimer disease model. Front Neurosci 2018;12:333. [doi: 10.3389/fnins.2018.00333] [Medline: 29872375]
56. Shah S, Mansour HM, Lucke-Wold B. Advances in stem cell therapy for Huntington's disease: a comprehensive literature review. Cells 2025 Jan 3;14(1):42. [doi: 10.3390/cells14010042] [Medline: 39791743]
57. Dong X. Current strategies for brain drug delivery. Theranostics 2018;8(6):1481-1493. [doi: 10.7150/thno.21254] [Medline: 29556336]
58. Bhunia S, Kolishetti N, Vashist A, Yndart Arias A, Brooks D, Nair M. Drug delivery to the brain: recent advances and unmet challenges. Pharmaceutics 2023 Nov 23;15(12):2658. [doi: 10.3390/pharmaceutics15122658] [Medline: 38139999]
59. Rafii MS, Baumann TL, Bakay RAE, et al. A phase1 study of stereotactic gene delivery of AAV2-NGF for Alzheimer's disease. Alzheimers Dement 2014 Sep;10(5):571-581. [doi: 10.1016/j.jalz.2013.09.004] [Medline: 24411134]
60. Cheng Y, Zhang S, Shang H. Latest advances on new promising molecular-based therapeutic approaches for Huntington's disease. J Transl Int Med 2024 Apr;12(2):134-147. [doi: 10.2478/jtim-2023-0142] [Medline: 38779119]
61. Ben Jehuda R, Shemer Y, Binah O. Genome editing in induced pluripotent stem cells using CRISPR/Cas9. Stem Cell Rev Rep 2018 Jun;14(3):323-336. [doi: 10.1007/s12015-018-9811-3] [Medline: 29623532]
62. Raffaele I, Cipriano GL, Anchesi I, Oddo S, Silvestro S. CRISPR/Cas9 and iPSC-based therapeutic approaches in Alzheimer's disease. Antioxidants (Basel) 2025 Jun 25;14(7):781. [doi: 10.3390/antiox14070781] [Medline: 40722885]
63. Yun Y, Ha Y. CRISPR/Cas9-mediated gene correction to understand ALS. Int J Mol Sci 2020 May 27;21(11):3801. [doi: 10.3390/ijms21113801] [Medline: 32471232]
64. Singh A, Smedley GD, Rose JG, et al. A high efficiency precision genome editing method with CRISPR in iPSCs. Sci Rep 2024;14(1):9933. [doi: 10.1038/s41598-024-60766-4] [Medline: 40722885]


## Abbreviations

**AADC:** aromatic L-amino acid decarboxylase
**AAV:** adeno-associated virus



XSL•FO
RenderX

**AD:** Alzheimer disease
**ALS:** amyotrophic lateral sclerosis
**Aβ:** amyloid-β
**BBB:** blood-brain barrier
**BDNF:** brain-derived neurotrophic factor
**Cas9:** CRISPR-associated protein 9
**CED:** convection-enhanced delivery
**CRISPR:** clustered regularly interspaced short palindromic repeats
**GDNF:** glial cell line–derived neurotrophic factor
**HD:** Huntington disease
**iPSC:** induced pluripotent stem cell
**MSC:** mesenchymal stem cell
**ND:** neurodegenerative disease
**PD:** Parkinson disease
**siRNA:** small interfering RNA
**TGCS:** total global citation score
**TLCS:** total local citation score
**WoSCC:** Web of Science Core Collection

*Edited by A Schwartz, M Balcarras; submitted 07.Sep.2025; peer-reviewed by M Yu, S Zhang; revised version received 27.Jan.2026; accepted 09.Feb.2026; published 09.Mar.2026.*

*Please cite as:*
*Xiang L, Xiao Y, Cai M, Qin J, Wang T, Xiang X, Ke J, Peng G*
*Global Research Trends and Hotspots in Gene Editing and Stem Cell Therapies for Neurodegenerative Diseases: Bibliometric and Visualization Analysis*
*Interact J Med Res 2026;15:e83709*
*URL: https://www.i-jmr.org/2026/1/e83709*
*doi:10.2196/83709*

© Lijun Xiang, Yun Xiao, Ming Cai, Jing Qin, Ting Wang, Xueming Xiang, Jun Ke, Ganlin Peng. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 9.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Mapping Digital Nudges and Recommender Systems for Obesity Prevention: Scoping Review

Sarah Forberger[1,2], PD, PhD; Lucia A Reisch[3], Prof Dr; Pieter Van Gorp[4], Prof Dr; Christoph Stahl[5], Dr.-Ing; Lara Christianson[2], MLS; Jihan Halimi[6], MSc; Karina Karolina De Santis[2], PhD; Chungwan Lo[2], MSc; Cassandra A Omane[2], MSc; Alejandra Loyola-Leyva[7], PhD; Laurent Malisoux[7], PhD; Tiziana de-Magistris[6], PhD; Torsten Bohn[7], PhD

[1]Leibniz Science Campus Digital Public Health (LSC DiPH), Bremen, Germany

[2]Leibniz Institute for Prevention Research and Epidemiology - BIPS, Achterstraße 30, Bremen, Germany

[3]El-Erian Institute for Behavioural Economics and Policy, Cambridge Judge Business School, Cambridge, United Kingdom

[4]Eindhoven University of Technology, Eindhoven, The Netherlands

[5]Luxembourg Institute of Science and Technology, Belvaux, Luxembourg

[6]Agro-Food Research and Tecnological Center of Aragon (CITA), Zaragoza, Spain

[7]Luxembourg Institute of Health, Strassen, Luxembourg

**Corresponding Author:**
Sarah Forberger, PD, PhD
Leibniz Science Campus Digital Public Health (LSC DiPH), Bremen, Germany

## Abstract

**Background:** Recommender systems are pivotal in organizing information to enhance noticeability, reduce overload, and streamline decision-making. They can be even more effective if combined with digital nudges. Digital nudging is a subtle approach that combines design, information, and interaction elements to create a choice architecture that can guide user behavior in digital environments. While promising in many fields, there is a notable gap in health promotion, particularly because digital nudges and recommender systems can encourage and support sustained healthier choices in nutrition, physical activity (PA), and sedentary behavior reduction to prevent overweight and obesity.

**Objective:** This scoping review addresses these gaps by exploring how digital nudges and recommender systems are used in obesity prevention.

**Methods:** We prospectively published the scoping review protocol and adhered to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines. Eligibility was defined using the PCC (Population, Concept, Context) framework. We searched 7 bibliographic databases (MEDLINE and PsycINFO via Ovid, Web of Science, CINAHL via EBSCO, Scopus, ACM Digital Library, and IEEE Xplore) up to October 2023. Following a 2-stage screening by independent reviewers, we selected 68 articles that included 94 user evaluations.

**Results:** Most articles (36/68, 53%) report on recommender systems focused on nutrition, with fewer (16/68, 23%) aiming to promote PA. Most studies on digital nudges (11/68, 16%) targeted nutrition-related nudges for shopping and meal selection (8/68, 2%). Articles address PA and sedentary behavior less frequently (3/68, 4%). Three out of 68 (4%) articles report on recommender systems in combination with games, and 2 out of 68 (3%) articles report on recommender systems and digital nudges. Approaches to item retrieval vary widely, with 31 out of 68 (46%) articles failing to describe their methods. In the scoping review, we found a discrepancy between the target group for which the system was developed and the group with which the evaluation was conducted. Sixty-eight evaluations report positive results, while 26 studies report mixed, negative, or no-difference results.

**Conclusions:** Integrating digital nudges and recommender systems might hold potential in overweight and obesity prevention by subtly encouraging healthier lifestyle choices. However, the heterogeneity in study designs, outcome measures, and reporting quality limits the comparability of findings and makes it difficult to draw robust conclusions about effectiveness. Future work should include detailed definitions, mechanism descriptions, broader geographic representation, and rigorous intervention testing and user evaluations to fully leverage these systems for improved health outcomes and to support sustainability and well-being objectives.

**International Registered Report Identifier (IRRID):** RR2-10.1136/bmjopen-2023-080644

***(Interact J Med Res 2026;15:e73151)*** doi:10.2196/73151



XSL•FO
RenderX

**KEYWORDS**

recommender systems; lifestyle; physical activity; sedentary behavior; obesity prevention; digital health; digital intervention; digital nudges

# Introduction

## Background

Recommender systems are tools designed to predict and suggest items, services, or content that users might like based on their preferences, behavior, or other data. For instance, recommender systems can suggest recipes informed by the user's past experiences and dietary preferences, such as vegan or vegetarian options. These preferences are discerned by analyzing and comparing the search histories of different users [1,2]. Furthermore, these systems can remind users to increase their physical activity (PA), integrate tasks into daily schedules based on personal calendars and to-do lists, calculate caloric expenditure, and propose recipes based on that expenditure, while also suggesting suitable exercise methods [3]. In this process, recommender systems use sophisticated algorithms to analyze extensive datasets. This approach facilitates tailored suggestions by applying diverse filtering techniques [4-6]. Therefore, recommender systems are considered digital interventions that leverage nudging mechanisms in the digital environment [7]. Digital nudging is an approach grounded in behavioral economics that utilizes user interface design elements to influence user choices in digital environments. These elements encompass, for example, graphic design, coloring, specific content, wording, ranking, and other features, collectively serving as subtle tools to guide individuals toward decisions or behaviors [8]. In other words, they provide a choice architecture in the digital space intended to guide individual decisions. The digital environment is where users make decisions on screens, such as tablets, websites, or mobile apps [9]. Nudging mechanisms, according to the taxonomy developed by Münscher et al [10] and reflected in the functionality of recommender systems, can be categorized into:

- Decision information: This refers to techniques that target the presentation of decision-relevant information without altering the decision options. This category focuses on how information is perceived, represented, and framed, influencing the decision-making process. It includes methods such as translating information to make it clearer, reframing choices to alter perspectives, simplifying complex data to enhance comprehension, and making critical information or social norms visible to guide decision-makers without changing the underlying content.
- Decision structure: This involves altering the arrangement and presentation of decision options to influence decision-making strategies and outcomes. This approach includes modifying the decision environment by setting defaults, adjusting the range or composition of available options, changing the effort required to select an option, and altering the potential consequences associated with each choice. The goal of these techniques is to guide decision-makers toward more desirable outcomes by organizing and structuring the decision space in a way that impacts their selection process.
- Decision assistance: This encompasses techniques aimed at supporting individuals in following through with their intentions to bridge the intention-behavior gap. This category includes interventions such as reminders to bring desired options to the forefront of attention and facilitating commitment to reinforce adherence to chosen behaviors. These techniques are designed to enhance self-regulation and assist decision-makers in maintaining their resolve, thereby increasing the likelihood of consistent and intended actions.

Recommender systems automatically select information that influences information noticeability [10]. They sort information based on preferences and prompt as suggestions to elicit a reaction from users. In doing so, recommender systems can reduce information overload, help rank information, and streamline the decision-making process [1,2].

While recommender systems inherently possess features such as information selection, simplification, and reminders akin to digital nudges, they are rarely included in the category of digital nudges in the field of health promotion. They are designed with these principles in mind. Interactions within these systems, such as visual feedback on daily step counts, calorie tracking, messages, and reminders, are infrequently developed with the explicit framework of digital nudges as the guiding principle. By effectively using design, information, and interaction elements, digital nudges aim to steer decision-making processes. Combining digital nudges and recommender systems could increase their individual effectiveness, as demonstrated in various studies [8,11-13]. However, only a small number of nudges with recommender system mechanisms have been investigated until now [7], and large gaps exist [14]. Furthermore, the literature is characterized by expanding research activity, volume, and heterogeneity [15].

However, given the effects that digital nudges and recommender systems can exert, this potential should be leveraged for obesity prevention, a crucial aspect of modern lifestyle interventions. According to Hummel et al [16], 62% of nudging treatments have statistically significant effects, with a median effect size of 21%, which varies by category and context. Defaults emerge as the most effective strategy (number of included studies: 21, median effect size: 50%; average effect size: 87%), while precommitment strategies are the least effective (number of included studies: 2, median effect size: 7%; average effect size: 7%). Additionally, digital nudging demonstrates comparable effectiveness and presents opportunities for individualization, allowing for more tailored approaches to influencing healthy behaviors. Obesity prevention has become an important part of public health prevention due to the drastic increase in obesity in the past decades, with adult obesity rates as high as 34% in some European countries and even 40% in the United States [17]. As the trends continue almost unencumbered, novel approaches are required to prevent and reduce obesity, and digital tools have aimed to contribute to this. Utilizing



XSL•FO
RenderX

recommender technologies and nudging strategies aids users in navigating the extensive and often heterogeneous data landscape. This approach is instrumental in identifying accurate and relevant information essential for making informed health-related decisions [18-20]. By using a personalized digital choice architecture, the way information, services, and choices are presented to users on websites and other online services can be altered. This approach can subtly encourage desired behaviors such as improved dietary habits to prevent weight gain and enhance metabolic health. Additionally, it can promote more physically active lifestyles and support maintaining healthy weights. Research into digital nudging analyzed how users perceive choices and make decisions in digital environments [7,8]. Combining this knowledge with recommender systems allows for presenting tailored suggestions within a choice architecture that naturally guides users toward desirable decisions. For example, combining healthy food items based on user preferences with visual progress bars or social comparison features (eg, "90% of users chose this option") could encourage users to follow recommendations [21,22]. General challenges reported by individuals who are gaining weight, such as making poor choices when tired or hungry, focusing on budget constraints rather than a balanced diet, lacking time to gather information on healthy food items or exercise, or staying motivated over a more extended period, can be addressed by both approaches.

Bergman et al [13] described various types of digital nudges based on digital nudge patterns, outcomes, context, evaluation, personalization, interconnectivity, and mode of delivery. However, the context was broadly labeled as "health," without specifically addressing the behavior targeted by these nudges. Similarly, Jesse and Jannach [7] developed a taxonomy for coding digital nudges, integrating the work of others by categorizing nudges and their mechanisms [10,22-24]. However, their work did not focus on the health context of the systems. A similar pattern can be seen in the literature on recommender systems. De Santis et al [25] identified 10 reviews from 2017 to 2023, covering 308 primary studies on recommender systems and nudges for obesity prevention, primarily focusing on technical system properties and strengths and weaknesses [1,26-35]. While health domains related to obesity prevention were mentioned, system implementation and evaluation with users were rarely described. De Croon et al [28], for example, analyzed the use of recommender systems in health applications aimed at laypersons and identified 26 nutrition-related recommender systems, including meal plans, recipes, restaurant suggestions, and fruits. Further, 24 studies were included under lifestyle, encompassing PA and the combination of nutrition and PA, as well as other lifestyle behaviors such as smoking. Similarly, Pincay et al [6] classified various studies analyzing recommender systems under "wellness," which covered diets, exercise routines, and care recommendations for children and older people. Pincay et al [6] did not distinctly separate nutrition from PA behaviors. Cai et al [2] identified 24 health domains. The health domain (specific health conditions, issues, or diseases) identified ranges from general healthy lifestyle promotion to specific health domains such as diabetes and obesity. Bondevik et al [1] analysis focused on technical aspects of food recommender systems, such as methods and techniques used within the recommender system (eg, filtering, personalization) data processing techniques (eg, data used for calculation similarities, data used for training the system), evaluation, and reproducibility without considering specific target groups, topics, or the integration of approaches such as digital nudges. Sun et al [36] offered a broader perspective, mapping health recommender systems based on characteristics, techniques, interfaces, and user involvement. However, health behaviors were broadly categorized under "lifestyle." No review has explicitly focused on nutrition, PA, and sedentary behavior (SB) reduction. Furthermore, given the diversity of contexts in which health behaviors occur, it is crucial to consider socioeconomic and geographical disparities, as these can significantly impact mechanisms, norms, acceptance, effectiveness, and, in the end, the generalizability of digital intervention strategies. More information is needed on current research approaches, target audiences, and employed methods and techniques to efficiently use recommender systems, digital nudges, or a combination of both for overweight and obesity prevention.

### Study Aims and Objectives

This paper addresses the gaps in mapping how digital nudges and recommender systems are currently used in research to target overweight and obesity by addressing nutrition, PA, and SB reduction.

The detailed objectives of this scoping review are as follows:

1. To identify digital nudges or recommender systems for overweight and obesity prevention that target diet, PA promotion, or SB reduction.
2. To map the digital nudges and recommender systems according to the target population, health behavior (diet, PA, or SB), and system classification (eg, mechanisms for developing recommendations, delivery channels, personalization, interconnection, and used combinations).

In doing so, our goal is to identify the current state of research, identify gaps, and identify the next steps in research. By understanding these gaps, researchers can better identify what needs to be addressed in the next steps and help avoid repeating approaches.

## Methods

### Study Design, Protocol, and Registration

A detailed description of the methods for this scoping review was previously published [37], and this scoping review follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [38]. We chose a scoping review because it allows us to map the key concepts and gaps within the area of interest. The work on this study began in July 2023, and the database searches were conducted in October 2023. Screening and study selection were conducted from October 2023 to March 2024, data extraction from April to October 2024, and data synthesis from November 2024 to February 2025.



XSL•FO
RenderX

## Ethical Considerations

No ethical approval was required for this review, as the data were obtained from publicly available materials.

## Eligibility

The eligibility criteria were based on the PCC (Population, Concept, Context) framework [39]. We included primary studies of any design (randomized or nonrandomized, with quantitative or qualitative data) (Table 1). We excluded all articles that tested the system using hypothetical or predefined datasets, focusing only on studies that tested it on a real-world sample.

**Table .** Eligibility for inclusion in the scoping review.

| PCC[a] | Inclusion criteria | Exclusion criteria |
|---|---|---|
| Population | Any human population groups (children or adults; healthy, at risk for chronic diseases, or clinical samples) | Studies that do not test the systems or nudges with humans, and studies that only use hypothetical or predefined datasets. |
| Concept | Digital nudges or recommender systems (stated in the title, abstract, or full text of a study) | Digital nudges or recommender systems not used, or studies that use digital nudges or rely on recommender systems but do not name them in the title, abstract, or full text |
| Context | Any geographical setting;<br>overweight and obesity prevention (eg, nutrition, food recipes, grocery stores, meal preparation, PA[b] promotion, SB[c] prevention) | Digital nudges or recommender systems used in fields, such as blockchain, finance, security, privacy, agriculture, service, and e-commerce rather than for overweight and obesity prevention |

[a]PCC: Population, Concept, Context.

[b]PA: physical activity.

[c]SB: sedentary behavior.

As PA and SB are not opposites but complementary, we included studies targeting PA and SB. PA is defined as any bodily movement generated by skeletal muscle contraction that increases energy expenditure above the resting metabolic rate and is characterized by its modality, frequency, intensity, duration, and context of practice [40]. In contrast, SB includes any waking behaviors characterized by an energy expenditure of ≤1.5 metabolic equivalents while in a sitting, reclining, or lying posture [41]. In general, this means that one can be both physically active (eg, meeting the WHO recommendations for PA) and highly sedentary (eg, accumulating a high amount of sitting time), which significantly influences the choice of intervention elements to be utilized, depending on the targeted specific behavior [42].

## Information Sources and Search Strategy

The search strategy was developed with the assistance of an information specialist, in collaboration with the involved experts (coauthors). We utilized the Joanna Briggs Institute (JBI) 3-step approach [43] to formulate the search strategy using keywords as well as index and MeSH terms [44]. The seven databases searched until October 2023 were MEDLINE and PsycINFO via Ovid, Web of Science, CINAHL via EBSCO, Scopus, ACM Digital Library, and IEEE Xplore. The full search strategy for Medline can be found in Multimedia Appendix 1.

## Study Selection

Two reviewers independently screened the records using predefined eligibility criteria based on the PCC framework (Table 1) in a 2-stage process. Titles and abstracts were screened in Rayyan, followed by a full-text screening in Covidence. Any disagreements were resolved through discussion at each stage between the 2 reviewers or through the involvement of a third reviewer.

## Data Charting and Data Items

A team of reviewers performed data extraction using a standardized, self-developed, and pilot-tested extraction form. After extraction, one researcher (SF) processed the data into predefined categories and verified the coding with a second team member (Table 2).



XSL•FO
RenderX

**Table .** Predefined categories.

| Code | Operationalization | Explanation |
|---|---|---|
| Country/region | • Country: as indicated in the publication, if not indicated in the publication, the country of the first author's affiliation | • High-income, upper-middle-income, lower-middle-income, and lower-income countries based on the World Bank country classifications by income level (2022 - 2023) |
| Recommender system classification [45] | • Methods: approach that the recommender system uses to perform the retrieval of items [6] | • Content-based filtering, where recommendations are made by identifying items similar to those previously liked by the user or other users with similar profiles [46,47].<br>• Collaborative filtering, which relies on similarities between users to recommend items based on the preferences of other users with similar profiles [48].<br>• Hybrid recommendation combines multiple techniques to overcome the limitations of single approaches [45]. An example is the deep learning approach to item retrieval in recommender systems that refer to the application of deep neural networks (deep learning) to filter relevant items (eg, products, films, songs) for a user from a vast pool of options. Deep learning is used in this context to uncover complex patterns within data, thereby enabling more precise and personalized recommendations.<br>• Knowledge-based recommendation is another approach based on user preferences and constraints [35]. Collecting the preferences of a user (preference elicitation) within the scope of a dialogue and then recommending items either (1) based on a predefined set of recommendation rules (constraints) or (2) using similarity metrics that help to identify items which are similar to the preferences of the user [35].<br>• Evolutionary approaches rely on optimization algorithms inspired by natural evolution to retrieve or rank items. These approaches are beneficial when the search space for items is vast, and the system needs to explore a variety of combinations to find the most optimal recommendations. |
| | • Techniques | • How the relationships between items are computed (eg, Rasch-tailoring, TOPIs[a], conversational agent, iALS[b] algorithm) [6]. |
| Age | • Children (0 - 14),<br>• 15 - 17, 18 - 40, 40 - 65, adults (18-65), ≥60, older adults (≥65), adults (18-65) and older adults (≥65), not reported | • Where possible, the mean age was used to report age data. If the mean age was not provided, age ranges were used, and the target group was assigned accordingly based on these age ranges. |
| Outcome | • Positive: only positive results are reported<br>• Negative: only negative results are reported<br>• Mixed: positive and negative results are reported<br>• No difference | • A qualitative assessment of the results was chosen to consolidate the extremely varied reporting of outcomes (such as significance levels, descriptive statistics, content analyses, and log data). |

[a]TOPI: Topic Models Over Points of Interest.

[b]iALS: Implicit Alternating Least Squares.



XSL•FO
RenderX

### Critical Appraisal

No evidence appraisal was conducted. In a scoping review, the primary aim is to map the existing literature and provide an overview of the breadth and scope of research on a particular topic rather than to assess the quality of the evidence. This type of review typically includes a wide range of studies, often without applying strict inclusion criteria related to methodological rigor. As a result, scoping reviews do not appraise evidence quality, focusing instead on identifying and categorizing available evidence.

### Overlap of Primary Studies Included in This Review and Other Reviews With Systematic Methodology

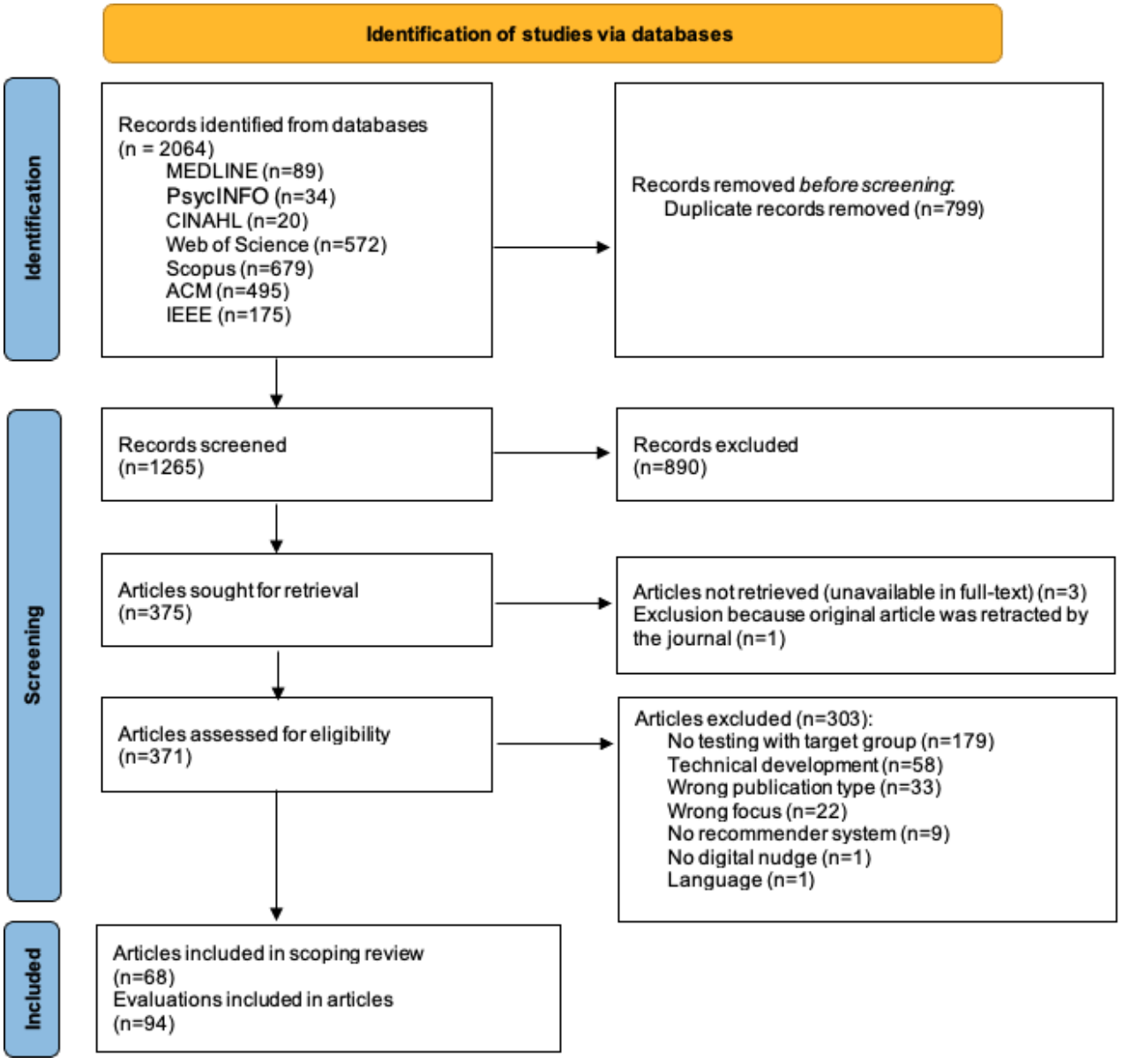
In addition to this scoping review of primary studies, we conducted a second scoping review of reviews identified in our search [49]. The scoping review of reviews included 10 reviews (n=8 systematic reviews and n=2 scoping reviews). The reviews addressed recommender systems for any human populations (n=6), described the systems (n=8), and focused on nutrition (n=6). The overlap between the primary studies included in the 10 reviews and those in this review was low. Specifically, of the 308 primary studies included in the 10 reviews, 10 primary studies were included in this scoping review.

## Results

### Characteristics of Included Studies

Of the 2064 records identified, 375 full-text articles were reviewed, and 68 articles covering 94 user evaluations were included (Figure 1). A detailed overview of the included studies is provided in Multimedia Appendix 2. Figure 2 provides an overview of the results.

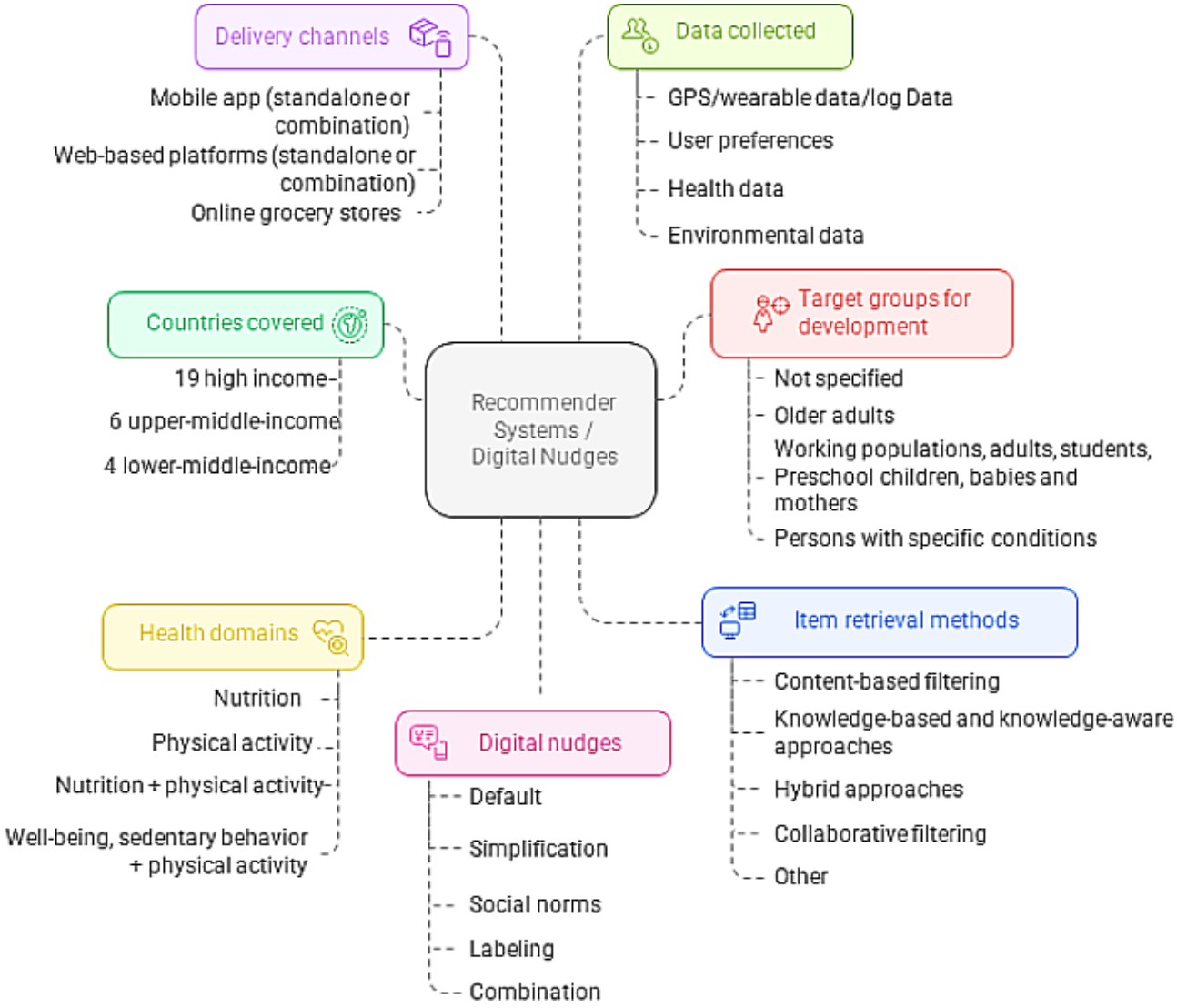
**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.






XSL•FO
RenderX

**Figure 2.** Overview of the results: study characteristics, behaviors addressed, item retrieval methods, and digital nudges used.




## Countries or Regions

Based on the World Bank's country classifications by income level (2022 - 2023), the articles were published from 19 high-income countries (65%), 6 upper-middle-income countries (21%), and 4 lower-middle-income countries (14%). The United States was the most represented among the high-income countries [50-58], followed by Germany and the Netherlands [59-73], Belgium [74-79], Italy [80-84], and Portugal [85-88].

The study overview table in Multimedia Appendix 2 provides a detailed overview of the countries and regions covered, as well as the study characteristics.

## Target Population Addressed for Whom the Digital Nudges or Recommender Systems Were Intended

Among the 68 articles, the majority did not specify the target group for which the digital nudges or recommender systems were intended (52/68, 76%). For those mentioning a target population, older adults (≥65 y) were the most frequently mentioned group (3/68, 4%, Table 3) [86-88]. All other target groups, such as mothers with babies [89], students [71], and persons with specific conditions, were each addressed once (Table 3). The addressed conditions in the articles included diabetes (2/68, 3%) [90,91], congenital [70] and coronary heart disease [57], and hypertensive patients [92] (except for diabetes, each mentioned once).



XSL•FO
RenderX

**Table .** Target groups for which the recommender systems or nudges have been developed.

| Target groups mentioned | N=68 | References |
|---|---|---|
| Not specified | 52 | —[a] |
| Target groups mentioned | | |
| Babies and mothers | 1 | [89] |
| Preschool children (3 - 6 y) | 1 | [85] |
| Students | 1 | [71] |
| Adults | 1 | [93] |
| Working population | 1 | [94] |
| Older adults (≥65 y) | 3 | [86-88] |
| Persons with specific conditions | | |
| Children with obesity | 1 | [95] |
| Adults with overweight and obesity | 1 | [53] |
| Diabetic patients | 2 | [91] |
| Hypertensive patients | 1 | [92] |
| Coronary heart disease patients | 1 | [57] |
| Patients with congenital heart disease | 1 | [70] |
| Insufficiently active people | 1 | [75] |

[a]Not applicable.

## Behaviors Addressed by the Digital Nudges and Recommender Systems

The articles predominantly focused on dietary behaviors, with 54 out of 68 (79%) articles covering nutrition. PA was addressed in 22 out of 68 (32%) articles. One article explicitly included SB, in combination with PA and mental health [94]. Fifteen out of 68 (22%) articles reported on PA.

Eight out of 68 (12%) articles explored nutrition and PA together [53,56,85,93,96-99] (Table 4).



XSL•FO
RenderX

**Table .** Behaviors addressed within the articles.

| Behavior | Articles, n (%) | References |
|---|---|---|
| Nutrition | 54 (78) | —[a] |
| PA[b] | 22 (32) | — |
| Behavior-specific nutrition | | |
| Nutrition (meal choice) | 30 (44) | RS[c]: [52,58,61-64,69,78,81-84,88,90,91,100-109]; mixed: RS + nudge: [63,110]; nudges: [65,69,108] |
| Nutrition (macronutrient composition) | 13 (19) | RS: [45,50,66,67,72,80,87,89,92,111,112]; nudges: [73] |
| Nutrition (shopping) | 7 (10) | RS: [51,79,113]; nudges: [55,59,77,114] |
| PA | | |
| PA | 15 (22) | RS: [54,57,60,68,74-76,86,115-117]; nudges: [70,71]; mixed (RS + game): [60] |
| Nutrition and PA | | |
| Nutrition (macronutrient composition), PA | 3 (4) | RS: [53,98]; mixed: RS + game: [93] |
| Nutrition (meal choice), PA | 3 (4) | RS: [56,96,99] |
| Nutrition (eating, drinking), PA, sleeping | 1 (1) | RS: [85] |
| Healthy living (diet, PA included) | 1 (1) | Mixed: RS + game: [95] |
| PA and mental well-being | | |
| PA, SB[d], physical and mental well-being | 1 (1) | Nudges: [94] |

[a]Not applicable.

[b]PA: physical activity.

[c]RS: recommender system; mixed: articles using RS and nudges or RS and games.

[d]SB: sedentary behavior.

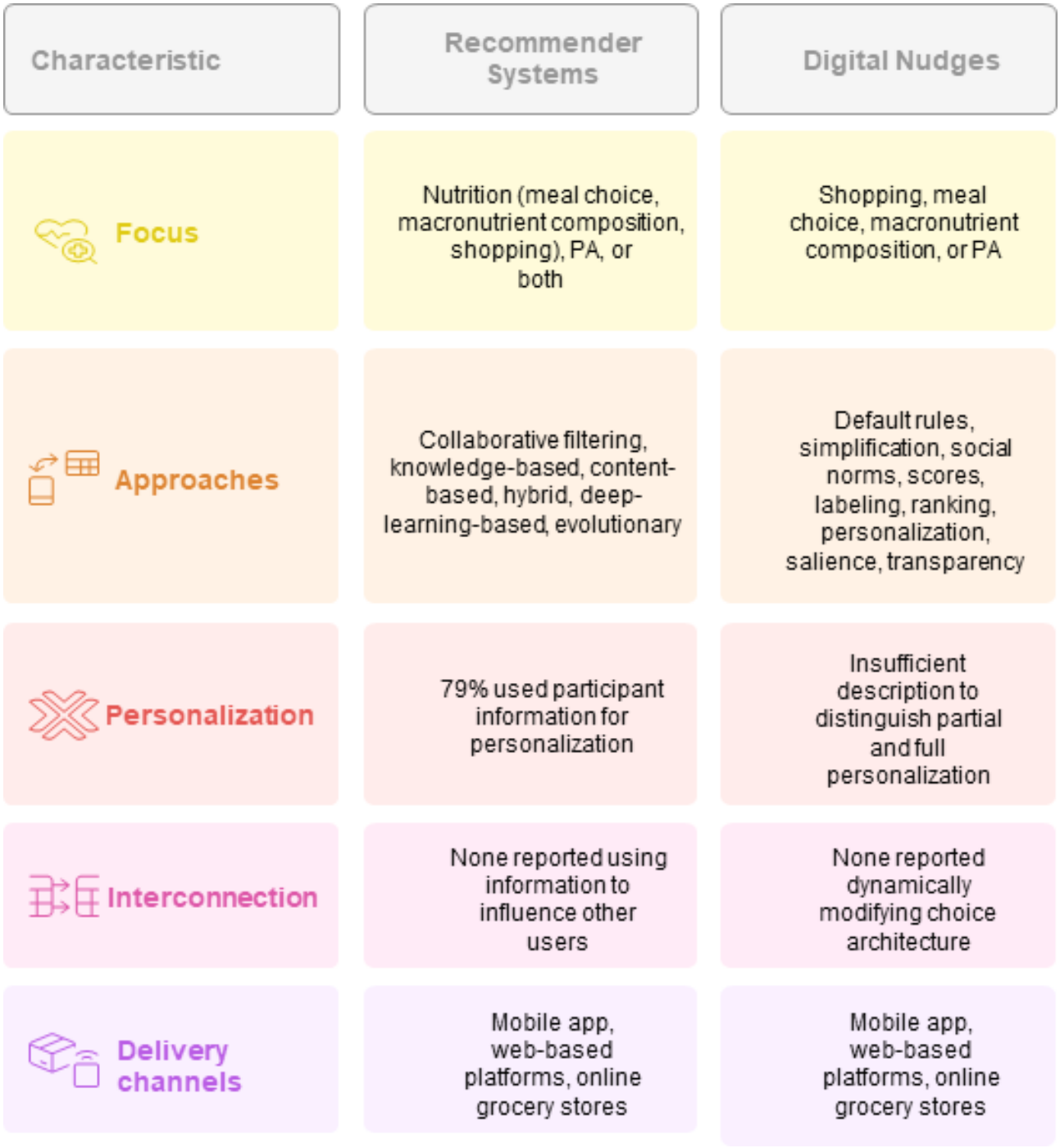
## System Characteristics (Recommender Systems, Nudges, Personalization, Interconnection, Delivery Channels)

Fifty-two out of 68 (76%) articles reported on recommender systems, 11 out of 68 (16%) articles focused on nudges [55,59,65,69-71,73,77,94,108,114] and 2 out of 68 (3%) articles discussed a combined system incorporating both a recommender system and a nudge [63,110]. Additionally, 3 out of 68 (4%) articles explored recommender systems in combination with games [60,93,95]. An overview of the characteristics can be found in Figure 3.



XSL•FO
RenderX

**Figure 3.** Characteristics: comparison between recommender systems and digital nudges. PA: physical activity.




## Recommender Systems

In the category of recommender systems (52/68, 76%), 36 out of 68 (53%) articles addressed recommender systems targeting nutrition (meal choice: 22/68, 32%; macronutrient composition: 11/68, 16%; shopping: 3/68, 4%). Additionally, 16 out of 68 (23%) articles reported recommender systems used to increase PA, 10 out of 68 (15%) articles purely targeting PA, and 5 out of 68 (7%) articles examined a combination of nutrition and PA, with 1 article also incorporating sleep behavior.

The approaches used by the recommender system to perform the retrieval of items [6] were collaborative filtering (2/68, 3%) [62,100], knowledge-based and knowledge-aware approaches (5/68, 7%) [64,67,84,89,90], content-based filtering (6/68, 9%) [64,66,79,87,92,105], hybrid approaches (6/68, 9%) [83,95,102,103,113,117], as well as, for example, deep-learning-based [107] and evolutionary [96] approaches (1/68 each) (definitions: Table 2) and various combinations of the systems. In 22 out of 68 (32%) articles, the methods used within the recommender systems were not further specified. In another 9 out of 68 (13%) articles, while the recommender system itself was not clearly defined, the accompanying techniques used were described, including rule-based reasoning, conversational agents, artificial intelligence (AI)–based personalization, or Rasch-based tailoring.



XSL•FO
RenderX

## Digital Nudges

Within the nudge category (11/68, 16%), 4 articles focused on shopping behavior [55,59,77,114], and 3 out of 68 (4%) articles addressed meal choice [65,69,108]. One article focused on addressing macronutrient composition [73]. Additionally, 3 out of 68 (4%) articles targeted PA [70,71,94], with 1 placing PA within a broader context, including SB and physical and mental health [94].

In articles reporting on digital nudges targeting nutrition behavior, various nudges were tested: default rules (eg, automatic assignment to a category, default setting of the food group, or recipe class), simplification (eg, simplified representation using images, pictograms, or the display of labels), and social norms (information on the behavior of what most people have done) [59]; the combined use of scores such as Nutri-Score and Eco-Score (graphics) and social norms [77]; social norms alone [55]; ranking and personalization [65]; or a combination of different approaches (defaults, social norms, warnings, defaults, and social cues) [114]. Combinations of salience (stimuli that capture the attention and influence decision-making, such as attractive or unattractive looking unhealthy food products), transparency (adding a message about what the nudge is intended to do), self-nudging (giving individuals the opportunity to adjust the nudge according to their preferences or decide whether they want it to be instantiated) [69], social norms, hedonic goal nudges and positional nudges [108], and attentional bias [73] were also explored.

In the context of PA studies, the digital nudges tested included providing information [70] and motivational prompts [71].

## Personalization and Interconnection

Forty-one out of 52 (79%) recommender articles reported using participant information for system personalization including partial personalization (eg, preferences, likes, GPS, health data, wearable data, survey data). Questionnaires for user preferences were commonly used, with some studies also incorporating wearable data for partial or full personalization. However, the descriptions were insufficient to clearly distinguish between partial personalization (where a study gathers user data—such as location, demographics, or user actions—to infer the potential influence of the nudge/recommender system on user behavior) and full personalization (where information is used to personalize the choice architecture of individual users dynamically).

None of the studies reported using this information to influence how information from other users affects the analyzed user behavior (partial interconnection). Similarly, none of the studies provided information on how the actions of one user, in turn, dynamically modify the choice architecture of other users (full interconnection).

## Delivery Channels

Nineteen out of 68 (28%) articles did not provide information on the mode of delivery or the devices used to access the systems. Of these, 14 out of 68 (21%) were recommender system studies and 5 out of 68 (7%) were digital nudge studies. The most commonly reported delivery method was through a mobile app (27/68, 40%), either as a stand-alone application or in combination with a desktop application, a wearable, or a game. Web-based platforms, either as a standalone approach or in combination with an app, were reported in 11 out of 68 (16%) articles [50,55,68,79,84,102-104,106,111,114,118]. Online grocery stores were used in 3 out of 68 (4%) studies [59,69,77]. One study did not fit into these categories because it was conducted in fast-food restaurants using online menu boards [108].

## Evaluations Reported

### *Population on Which the Digital Nudges/Recommender Systems Were Tested*

In the 68 included papers, 94 evaluations were reported. Most systems were evaluated with individuals aged 18 to 40 years (28/94, 26%), followed by adults aged 18 to 65 years (17/94, 18%). Further evaluation groups included combinations of adults aged 18 to 65 years and ≥60, 18 to 65 and ≥65, 18 to 40 and ≥65, and 40 to 65 years (altogether 12/94, 13%). Other individuals included were those aged 15 to 17 years, alone and in combination with other age groups (2/94, 2%), and those aged 0 to 14 years (2/94, 2%). In 23 out of 94 (24%) evaluations, the age range of the evaluation group was not reported. Furthermore, 53 out of 94 (78%) evaluations reported the gender of the evaluation group, with 51 (54%) reporting mixed gender and 2 (2%) reporting exclusively on women [89,93].

The sample sizes used for the evaluation varied substantially (Table 5). In 5 evaluations, no sample size was provided.



XSL•FO
RenderX

**Table .** Overview of the participants in the evaluations.

| Sample size | Evaluations[a] | | | | | | |
|---|---|---|---|---|---|---|---|
| | RS[b] | | | Digital nudges | | | RS + nudges, RS + game |
| | Nutrition | PA[c] | Mixed | Nutrition | PA | Mixed | |
| 0 - 10 | [87,89,92] | [54,57,68,86] | [56,95,97,98] | —[d] | — | — | — |
| 11 - 20 | [45,54,67,79,87,106,107] | [116] | [56] | [65] | — | — | RS + game, mixed behavior: [93,95] |
| 21 - 30 | [50,88-91,101,109,111] | [75,76,115] | — | — | — | — | — |
| 31 - 40 | [66] | [74,117] | [85] | — | — | — | — |
| 41 - 50 | [82,102,105] | — | — | — | — | [94] | — |
| 51 - 100 | [58,72,78,100,113] | — | [53,96] | [55,114] | [70,71] | — | RS + game, mixed behavior: [95] |
| 101 - 200 | [51,61,64,84,104] | — | [99] | [108,114] | — | — | — |
| 201 - 300 | [58,63,109] | — | — | [59,73] | — | — | — |
| 301- | [62,64,83,112] | [60] | — | [69,77] | — | — | RS + nudges, nutrition: [110] |
| No reporting | [52,72,103] | — | — | — | — | — | — |

[a]Some articles contain several evaluations.

[b]RS: recommender system.

[c]PA: physical activity.

[d]Not available.

Twenty-five out of 94 (27%) evaluations reported intervention time. The shortest intervention time reported was 7 days [92,97], while the longest was 3 years [95].

Twenty-five out of 94 (27%) evaluations reported an evaluation design. These designs included randomized controlled trials (7/94, 7%) [53,69-71,75,77,94], 2×2 between-subject designs (4/94, 4%) [62-64,104], online experiments (4/94, 4%) [55,59,73,109], single-case experiments [56], 2×2 mixed research designs [64,72], and quasi-experimental designs [57] (2/94, 2% each). Further designs were, for example, pre-post surveys [50], field experiments [59], longitudinal between-subject designs [76], effect differences between groups and within subjects [66], and 2×2 within-subject experiments [65].

The instruments used for evaluation ranged from questionnaires [51,67,81,87-89,95] and online surveys [60,78,91,96,110,113], focus groups [78], and ratings [61,82,84,95] to combinations of these methods with task performance, observations, gameplay monitoring, or user data.

### *Evaluation Results*

This scoping review includes 94 evaluations, of which 68 report positive results. There are 26 studies with mixed, negative, or no-difference results. In the domain of recommender systems related to nutrition, there were 8 studies with mixed results [45,67,79,84,87,89,92,109], 3 with negative outcomes [61] (2 evaluations in [64]), and 1 showing no difference [52], totaling 12 studies. For recommender systems targeting PA, 4 studies reported mixed outcomes [54,68,76,116]. In the case of recommender systems combined with nudges, 1 study reported mixed results [110]. Two studies reported mixed outcomes when combining recommender systems and games [60,95]. Studies involving nudges showed mixed results across 7 evaluations [55,59,70,77,94,108,118], with 1 showing no difference [71]. It is important to note that further comparison is not feasible due to the extreme heterogeneity of approaches, methods used, outcomes, and research questions, ranging from, for example, user and system evaluation, user satisfaction evaluation, system or chat performance, choice satisfaction, plausibility tests, or intervention validations across these studies (Multimedia Appendix 2).

## Discussion

### Principal Findings

Recommender systems help users avoid information overload and aid in decision-making by using filtering techniques to suggest relevant content. These systems typically present users with a curated list of items tailored to their preferences or needs, simplifying the process of finding and evaluating relevant information. By doing so, they streamline decision-making and make information searching more efficient [10]. These approaches can be classified as nudges, allowing recommender systems to be grouped under the category of digital nudges. This scoping review aimed to identify digital nudges or recommender systems for overweight and obesity prevention targeting diet, PA promotion, or SB reduction and to map the



XSL•FO
RenderX

digital nudges and recommender systems according to the target population, health behavior (diet, PA, or SB), and system classification (eg, mechanisms for developing recommendations, delivery channels, personalization, interconnection, and used combinations).

Recommender systems identified in this work predominantly focused on dietary behavior, with less emphasis on promoting PA. Additionally, the studies found that reporting on digital nudges primarily focused on nutrition, influencing shopping behavior and meal choices. Other health behaviors, such as PA and SB, are rarely reported. Only one study reported a combination of recommender systems and nudges. With this overview, the scoping review aligns with other reviews that reported that, in health promotion, the combination of recommender systems and nudges is underused. Jesse and Jannach [7] identified 9 domains in 18 articles that combined digital nudges and recommender systems. The domains were route planning, e-commerce, internet services, and social media, with route planning and e-commerce being the most popular ones.

This scoping review underscores the imbalance in the geographical distribution of research on recommender systems and digital nudges aimed at tackling overweight and obesity prevention [13]. High-income countries dominate the field, with the United States, Germany, the Netherlands, Belgium, Italy, and Portugal among the most represented. This distribution indicates that most research is focused in already technologically advanced regions, leaving middle- and low-income countries underrepresented. This discrepancy reflects broader socioeconomic disparities, including access to funding, infrastructure, and technology necessary to conduct such research. Expanding research efforts to include middle- and low-income countries would help to understand how digital interventions function across diverse socioeconomic contexts. Such inclusivity is crucial, as health behaviors such as diet and PA are global concerns, and solutions tailored to high-income settings may not be as effective in other contexts. In addition, overweight and obesity rates are high in many low- and middle-income countries [17]. A focus on a few countries hinders the adaptation of the identified digital nudges and recommender systems to other cultural, normative, and geographical contexts, thereby limiting the generalizability of the findings [1].

Regarding the second objective—classifying these systems based on the target population, behavior, personalization, and delivery—it was found that all recommender systems reported in the scoping review offered personalized recommendations. In this context, the findings of the scoping review differ from those reported in the literature. Bondevik et al [1], in their review of food recommender systems, state that 50.7% of the included recommender systems are not personalized. This discrepancy may be attributed to the slightly different scope of the scoping review. The present review only includes studies that explicitly evaluated the systems on users, which may result in a higher rate of personalization.

While many articles on recommender systems focus primarily on technical features [1,2,28,36], a central challenge identified in this scoping review is the lack of a clear and consistent definition of recommender systems, particularly in their application to health behaviors. Defining what constitutes an "actual system" can be ambiguous. Many studies describe systems that support users in selecting from a large pool of options without explicitly naming the system's mechanisms for item retrieval. This difficulty in identifying recommender systems stems from insufficient descriptions in the literature, limiting the reproducibility and understanding of how these systems function. Future research should aim for more transparent reporting, providing clear system descriptions and code access [1]. This would also enable research to transition the tested systems into user studies—a gap also identified in this work. Currently, systems are developed and evaluated for accuracy, user-friendliness, and applicability [1,2]. However, real-world intervention and implementation evaluation with repeated measurement over a longer time and larger sample sizes are needed [2]. Also, studies testing the same recommender systems in different target groups are missing. Both would increase the generalizability of the results. This is consistent with the result that 54 articles did not mention the intended target group. The evaluations, however, were reported very heterogeneously, often with small sample sizes of adults (18-65) or younger persons (18-40). Further studies are needed to combine explicitly and test recommender systems and nudging interventions in robust user studies (including trials, mixed methods approaches, and interdisciplinary research) to identify approaches with the potential for a long-term positive impact also requested by El Majjodi et al [119].

Despite the potential for digital nudges to enhance adherence to recommender systems, this scoping review reveals that many systems fail to capitalize fully on nudges, which aligns with Jesse et al [7], who reported that 69 of 87 nudging mechanisms identified in the literature have not yet been tested. Visualizations of behaviors, such as calorie consumption or activity levels, are commonly used; however, they are often disconnected from established nudging strategies. More explicitly, integrating nudges into these systems allows users to be more effectively influenced toward healthier behaviors. For instance, intuitive graphics, personalized comparisons, and progress summaries can serve as subtle nudges that reinforce positive behavior change. However, these approaches often lack a connection to digital nudging research [7,120]. By leveraging these approaches more purposefully within design feedback tools, these systems could achieve greater user engagement and more impactful behavior change [121]. Future research should focus on developing methods that incorporate these insights to maximize the impact of recommender systems in health interventions. Furthermore, beyond health behaviors, recommender systems have a broad potential to contribute to sustainability and well-being goals, as noted in recent studies [122,123]. For example, recommender systems could help optimize energy consumption in light of sustainability and environmentally friendly lifestyle choices [124,125]. More research is needed to harness this potential to fully integrate behavioral insights into recommender systems and increase the selection share of sustainable and healthy items and to support behavior [126-129]. This must be accompanied by qualitative research addressing accessibility questions, balance (eg, healthy



XSL•FO
RenderX

choices vs occasional indulgences), when the systems become intrusive, and how to balance taste preferences with health and environmental considerations—this may require perhaps options to preselect the main foci. El Majjodi et al [110] highlighted a crucial point that has often been overlooked in many system developments: the inherent capacity of these systems to learn. This could result in learning effects that deviate in undesirable directions. For example, systems might learn users' tastes and recipe preferences, which, over time, may no longer align with healthy options. As the systems evolve based on the prioritization of past preferences, their parameters shift accordingly. Currently, there is no research or discussion on how such developments are monitored or compared against the initial preferences entered by users and what the user accepts.

Traditional evaluations of recommender systems have primarily focused on prediction quality and user satisfaction, utilizing a combination of online and offline evaluations, as also found in the included studies [1]. However, as Atas et al [130,131] point out, there are other critical dimensions to consider, such as how users perceive the recommended items or the visibility of certain items in recommendation lists. Atas et al [130] argue that preferences are unstable and subject to change during the recommendation process. These factors directly influence user behavior and the effectiveness of the systems in guiding healthier choices. Future evaluations should adopt a more holistic approach that includes these additional metrics, ensuring that recommender systems are accurate and effective in shaping user behavior toward positive outcomes.

### Implications for Research

This paper presents the current state of research on digital nudges and recommender systems for obesity prevention, highlighting important gaps in evidence and practice. While both approaches might be effective, the current evidence is mixed. Provided that methodology and evaluation standards are harmonized, and stronger interdisciplinary collaboration between intervention experts, user experience scientists, ethicists, and implementation researchers is established, these technologies can be applied in a more targeted and sustainable manner. This would also enable comparative effectiveness trials to provide stronger evidence on the effectiveness of these approaches in this field. Knowing about these gaps is crucial because it enables researchers to target their efforts where they can make the greatest impact and helps avoid repeating ineffective approaches or overlooking promising new directions. This scoping review identified various points that should be targeted in the future:

First, there is a pressing need to harmonize the definitions of recommender systems, interventions, digital nudges, and evaluation methodologies used in obesity prevention research. Clearer definitions and standards for system properties, targets, and evaluation foci will make studies more comparable and cumulative [126,132].

Second, advancing this field will require genuine interdisciplinarity, integrating research and expertise from intervention development, user experience (UX) science, ethics and data governance, implementation, and behavioral science. The collaboration can enhance the design and effectiveness of digital nudges and recommender systems and ensure that interventions are contextually appropriate, user-centered, and scalable in real-world settings. Combining intervention development and UX research addresses usability, engagement, satisfaction, and adaptation of systems for diverse populations based on rigorous behavioral theory and empirical testing. Implementation science frameworks help to evaluate how well these interventions are adopted, reach their intended audiences, and sustain their impact beyond pilot trials. Insights from behavioral science, such as optimal design, timing, and frequency of triggers, reminders, labels, and nudges, are indispensable for the technical aspects. By systematically connecting these disciplines, future research can address not only technological and efficacy questions but also barriers to adoption, ethical data use, and the real-world conditions necessary for population-level obesity prevention [132].

Third, current systems frequently use static user preferences but rarely leverage the learning capability and dynamic adjustment demonstrated in commercial products. Real-time personalization, adaptation, and interconnection among users remain underutilized in research interventions, though they are increasingly utilized commercially. A critical discussion must also address how commercial products that already integrate these advanced features could be leveraged for research purposes without compromising personal data ownership, trust, transparency, and adherence to ethical standards. Balancing the use of commercially developed technologies with robust protections for user privacy and data governance is crucial to maintaining public trust and meeting high ethical standards in health research and intervention.

Fourth, data governance, privacy, and user ownership of data were largely absent from the reviewed studies. Yet, these are central to building trust in future health technologies and ensuring ethical and legal compliance in the dissemination of expanded systems. This includes establishing clear standards and best practices for the transparent use of data, informed consent, secure data storage, and providing users with meaningful control over their personal information. Collaborations among ethicists, legal experts, technology developers, and end-users are essential to co-create solutions that foster trust, comply with evolving regulations, and enable responsible innovation and research. It is essential to highlight the positive aspects of the systems but also to address areas of risk, such as developers' underlying assumptions, backward learning, and echo chambers. Such discussions are rarely included in current work.

Fifth, the scoping review predominantly focused on individual-level interventions. However, individual prevention alone cannot realize its full potential without supportive environments and population-level policy frameworks that enable healthy choices and prevent weight gain at scale. Population-level approaches are essential because they create the social, economic, and physical contexts necessary for sustainable behavior change. Integrating individual-level digital interventions, such as nudges and recommender systems, within broader population health policy approaches might be a long-term goal. This integration requires developing and evaluating multilevel interventions that consider policy, environmental, community, and health care system factors



XSL•FO
RenderX

alongside personalized behavior change tools. By bridging individual and population approaches, research can better address the complex determinants of obesity and enable impactful, equitable prevention strategies at scale. This approach is a very complex one and needs, as a starting point, evidence-based and effective interventions.

## Limitations

Several limitations emerged during this review. First, many studies were excluded because they did not test their recommender systems in real-world settings, focusing instead on hypothetical scenarios or predefined datasets. This exclusion limits the generalizability of the findings, as it is unclear how these systems would perform in practical applications. Additionally, we excluded studies that did not label their nudging approach as digital nudges. This decision likely led to the exclusion of some studies that might have been relevant to our analysis. For instance, studies focusing on shopping behaviors or the use of labels in online stores may have used digital nudging techniques or interventions in the digital environment but were not explicitly categorized as such. Consequently, these studies might have been overlooked in our scoping review. This limitation underscores the importance of how nudging interventions are described and classified in research, as it impacts the comprehensiveness of evaluations in the field. A potential limitation of our scoping review is the identification of SB within the broader context of PA. In our scoping review, we directly searched for SB to identify and include studies addressing SB. We only found one study that specifically addressed SB. While SB can be considered part of a continuum of movement behavior, some studies may have indirectly included it without explicitly mentioning or assessing it as a distinct factor. Many studies likely focused solely on moderate to vigorous PA and consequently overlooked SB, one of the continuum's extremities. This oversight is significant in light of ongoing debates regarding whether SB should be viewed as part of the movement behavior spectrum or as a separate entity [41,133]. Many publications fail to address SB independently, which can lead to a lack of emphasis on understanding its uniqueness and implications. Therefore, future research should explicitly consider and differentiate SB from other forms of PA to ensure a comprehensive understanding of movement behaviors and their impacts.

Another limitation is that we did not conduct a formal quality assessment. Scoping reviews are designed to map the extent and nature of available research rather than critically appraise the quality of the evidence, as would be the focus of a systematic review. Given the significant heterogeneity in methodologies, outcomes, and study designs, applying a uniform quality appraisal would not have been feasible without the risk of oversimplifying the diverse research landscape. Consequently, while this scoping review highlights the breadth of existing literature and its current gaps, the lack of quality assessment may affect the generalizability of the findings.

In line with the nature of scoping reviews as outlined above, this scoping review aims to provide a comprehensive overview of the existing literature in a particular research area, without focusing on specific outcome effects. The diverse range of topics and breadth of evidence precluded a detailed methodological quality assessment, which, as stated, might not have been feasible or appropriate given the exploratory aims of this scoping review. The broad scope and diversity of included studies make it challenging to report effect sizes or significances, leading to a qualitative categorization of outcomes as positive, negative, or mixed. This categorization choice is aligned with the scoping review's objective to include studies presenting both qualitative and quantitative findings, further emphasizing potential limitations in generalizability and effect estimation. The primary focus remains on identifying key concepts and knowledge gaps rather than evaluating the effectiveness of specific interventions, thus necessitating caution when applying these findings to specific contexts or populations.

Additionally, the review revealed that many studies do not report essential system details, such as how they personalize recommendations or use user data, making it challenging to assess the full scope of their effectiveness. Future research should prioritize real-world testing and comprehensively describe system design, personalization, and interconnection mechanisms to enhance transparency and reproducibility.

A final consideration concerns the timing of our database search, which was completed in October 2023 and may, therefore, not fully reflect the most recent developments in the field, particularly regarding AI and machine learning. While generative AI and large language models have attracted increasing attention during this period, our search strategy focused on applications evaluated in human populations rather than on studies using predefined datasets. Consequently, some of the most recent innovations may not yet be represented among the studies identified. Future reviews could incorporate emerging research on generative AI and large language models to capture these dynamic developments more fully.

## Conclusion

Combining digital nudges and recommender systems can play a crucial role in overweight and obesity prevention by encouraging healthier lifestyle choices and subtly, nonintrusively guiding users toward better nutrition, PA, and SB habits. However, their effectiveness and, in particular, which mechanisms may be helpful have not yet been fully explored. More details on definitions and mechanisms, as well as more inclusive geographical representation, intervention and implementation tests, and user evaluation studies, are essential for advancing the field. Previous research has often focused on individual-level interventions; however, for a tangible impact at the population level, integration into multilevel approaches, including policy and environmental strategies, is essential. The personalized, dynamic adjustment of nudges in combination with recommender systems, alongside assurances of data privacy and user trust, offers promising avenues for future investigation. This underlines that the contribution of this review is not merely descriptive but emphasizes the advancement of interdisciplinary methodology, evaluation frameworks, and scalability of digital interventions as critical priorities for ongoing research. By addressing these challenges, future studies can better harness the potential of these systems to improve health outcomes and contribute to broader goals such as sustainability and well-being.



XSL•FO
RenderX

## Acknowledgments

HealthyW8 Working Group:

LIH: Torsten Bohn, Farhad Vahid, Laurent Malisoux, Yvan Devaux, Manon Gantenbein, Michel Vaillant, Jonathan Turner, Mahesh Desai, Adriana Voicu

LIST: Christoph Stahl, Yannick Naudet, Jorge Ribeiro

NIUM: Alberto Noronha, Adam Selamnia

DFKI: Serge Autexier, Anke Koenigschulte

Virtech: Roumen Nikolov, Vladislav Jivkov, Alexandre Chikalanov

BIPS: Sarah Forberger

SPORA: Mireia Faucha, Edo Bazzaco

CREDA: Zein Kallas, Amelia Sarroca, Djamel Rahmani

USG: Maria Giovanna Ornati

CNR: Arianna D'Ulizia, Alessia D'Andrea

CITA: Tiziana de-Magistris

UEV: Elsa Sousa De Lamy

IDISBA: Josep Antoni Tur Mari, Cristina Bouzas

AOUBO: Giuseppe Tarantino, Michele Stecci

DTU: Rikke Andersen, Gitte Ravn-Haren

UT: Ying Wang

UC: Daniela Rodrigues

RCNE: Yoanna Ivanova, Boyko Doychinov

TU/e: Pieter Van Gorp, Astrid Kemperman

MEDEA: Pietro Dionisio

EADS: George-Mihael Manea

ENHA: Joost Wesseling, Chrysa Makridi

KNEIA: Cristina Barragan, Ciro Avolio

EFAD: Marianna Kalliostra, Ezgi Kolay, Katarzyna Janiszewska

## Funding

This study was conducted within the HealthyW8 project, which was funded by a European Union grant from the European Health and Digital Executive Agency (HaDEA), project number 1010806. The funder did not influence the content of this study.

## Authors' Contributions

SF conceptualized the study. SF, LAR, PVG, ALL, LM, KKDS, TDM, and TB developed the methodology and critically reviewed the manuscript. LC conducted the literature searches. SF, CS, TDM, JH, CL, CAO, and LM screened the studies. SF, CL, and CAO extracted and processed the data. SF analyzed the data, visualized the results, and wrote the first draft of the manuscript. All authors reviewed and edited the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search query.



XSL•FO
RenderX

[DOCX File, 15 KB - i-jmr_v15i1e73151_app1.docx ]

Multimedia Appendix 2
Table S1: Study characteristics.
[DOCX File, 73 KB - i-jmr_v15i1e73151_app2.docx ]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 103 KB - i-jmr_v15i1e73151_app3.pdf ]

## References


1. Bondevik JN, Bennin KE, Babur Ö, Ersch C. A systematic review on food recommender systems. Expert Syst Appl 2024 Mar;238:122166. [doi: 10.1016/j.eswa.2023.122166]
2. Cai Y, Yu F, Kumar M, Gladney R, Mostafa J. Health recommender systems development, usage, and evaluation from 2010 to 2022: a scoping review. Int J Environ Res Public Health 2022 Nov 16;19(22):15115. [doi: 10.3390/ijerph192215115] [Medline: 36429832]
3. Raza S, Rahman M, Kamawal S, et al. A comprehensive review of recommender systems: transitioning from theory to practice. Comput Sci Rev 2026 Feb;59:100849. [doi: 10.1016/j.cosrev.2025.100849]
4. Goldberg D, Nichols D, Oki BM, Terry D. Using collaborative filtering to weave an information tapestry. Commun ACM 1992 Dec;35(12):61-70. [doi: 10.1145/138859.138867]
5. Li Y, Lu L, Xuefeng L. A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in e-commerce. Expert Syst Appl 2005 Jan;28(1):67-77. [doi: 10.1016/j.eswa.2004.08.013]
6. Pincay J, Teran L, Portmann E. Health recommender systems: a state-of-the-art review. Presented at: 2019 Sixth International Conference on eDemocracy & eGovernment (ICEDEG); Apr 24-26, 2019. [doi: 10.1109/ICEDEG.2019.8734362]
7. Jesse M, Jannach D. Digital nudging with recommender systems: survey and future directions. Comput Hum Behav Rep 2021;3:100052. [doi: 10.1016/j.chbr.2020.100052]
8. Mirsch T, Lehrer C, Jung R, editors. Digital nudging: altering user behavior in digital environments. Presented at: Proceedings der 13 Internationalen Tagung Wirtschaftsinformatik; Feb 12-15, 2017; St Gallen, Switzerland p. 634-648 URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1076&context=wi2017 [accessed 2026-03-09]
9. Mitchell EG, Heitkemper EM, Burgermaster M, et al. From reflection to action: combining machine learning with expert knowledge for nutrition goal recommendations. Proc SIGCHI Conf Hum Factor Comput Syst 2021 May;2021:206. [doi: 10.1145/3411764.3445555] [Medline: 35514864]
10. Münscher R, Vetter M, Scheuerle T. A review and taxonomy of choice architecture techniques. Behav Decis Mak 2015 Aug;29(5):511-524. [doi: 10.1002/bdm.1897]
11. Weinmann M, Schneider C, Brocke JV. Digital nudging. Bus Inf Syst Eng 2016 Dec;58(6):433-436. [doi: 10.1007/s12599-016-0453-1]
12. Dalecke S, Karlsen R. Designing dynamic and personalized nudges. Presented at: Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics (WIMS 2020); Jun 30 to Jul 3, 2020. [doi: 10.1145/3405962.3405975]
13. Bergram K, Djokovic M, Bezençon V, Holzer A. The digital landscape of nudging: a systematic literature review of empirical research on digital nudges. Presented at: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; Apr 29 to May 5, 2022. [doi: 10.1145/3491102.3517638]
14. Decrinis L, Reisch LA. Nudging employees for corporate sustainability: a systematic evidence map. In: Sunstein CR, editor. Research Handbook on Nudges and Society: Edward Elgar Publishing; 2023:152-173. [doi: 10.4337/9781035303038.00017]
15. Valta M, Maier C. Digital nudging: a systematic literature review, taxonomy, and future research directions. ACM SIGMIS Database 2025 Feb;56(1):101-125. [doi: 10.1145/3715966.3715973]
16. Hummel D, Maedche A. How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. J Behav Exp Econ 2019 Jun;80:47-58. [doi: 10.1016/j.socec.2019.03.005]
17. World obesity atlas 2024. : World Obesity Federation; 2024 URL: https://data.worldobesity.org/publications/WOF-Obesity-Atlas-v7.pdf [accessed 2026-06-09]
18. Bai X, Wang M, Lee I, Yang Z, Kong X, Xia F. Scientific paper recommendation: a survey. IEEE Access 2019;7:9324-9339. [doi: 10.1109/ACCESS.2018.2890388]
19. Feng J, Wei C, Jiang S, Zhou M, Cao X, Pogue B, et al. MRI guided wearable near infrared spectral tomography: simulation study. Presented at: SPIE Photonics West – BiOS: Biomedical Optics Symposium; Mar 6-11, 2021. [doi: 10.1117/12.2578356]
20. Trang Tran TN, Atas M, Felfernig A, Stettinger M. An overview of recommender systems in the healthy food domain. J Intell Inf Syst 2018;50(3):501-526. [doi: 10.1007/s10844-017-0469-0]
21. Castmo M, Persson R. The alliance between digital nudging & persuasive design [Master's Thesis]. : Lund University, School of Economics and Management; 2018 URL: https://www.lunduniversity.lu.se/lup/publication/8950219 [accessed 2026-03-09]




XSL•FO
RenderX

22. Caraban A, Karapanos E, Gonçalves D. 23 ways to nudge: a review of technology-mediated nudging in human-computer interaction. Presented at: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19); May 4-9, 2019. [doi: 10.1145/3290605.3300733]
23. Dolan P, Hallsworth M, Halpern D, King D, Metcalfe R, Vlaev I. Influencing behaviour: the mindspace way. J Econ Psychol 2012 Feb;33(1):264-277. [doi: 10.1016/j.joep.2011.10.009]
24. Sunstein CR. The council of psychological advisers. Annu Rev Psychol 2016;67:713-737. [doi: 10.1146/annurev-psych-081914-124745] [Medline: 26393867]
25. De Santis KK, Stiens L, Christianson L, Forberger S. Recommender systems for obesity prevention: scoping review of reviews. SAGE Open Med 2025;13:20503121251348374. [doi: 10.1177/20503121251348374] [Medline: 40546930]
26. Abhari S, Safdari R, Azadbakht L, et al. A systematic review of nutrition recommendation systems: with focus on technical aspects. J Biomed Phys Eng 2019 Dec;9(6):591-602. [doi: 10.31661/jbpe.v0i0.1248] [Medline: 32039089]
27. Azmi AK, Abdullah N, Emran NA. A recommender system model for improving elderly well-being: a systematic literature review. Int J Advance Soft Compu Appl 2019;11(2) [FREE Full text]
28. De Croon R, Van Houdt L, Htun NN, Štiglic G, Vanden Abeele V, Verbert K. Health recommender systems: systematic review. J Med Internet Res 2021 Jun 29;23(6):e18035. [doi: 10.2196/18035] [Medline: 34185014]
29. Ferretto LR, Cervi CR, de Marchi ACB. Recommender systems in mobile apps for health a systematic review. Presented at: 2017 12th Iberian Conference on Information Systems and Technologies (CISTI); Jun 21-24, 2017. [doi: 10.23919/CISTI.2017.7975743]
30. Kirk D, Catal C, Tekinerdogan B. Precision nutrition: a systematic literature review. Comput Biol Med 2021 Jun;133:104365. [doi: 10.1016/j.compbiomed.2021.104365] [Medline: 33866251]
31. Orue-Saiz I, Kazarez M, Mendez-Zorrilla A. Systematic review of nutritional recommendation systems. Appl Sci (Basel) 2021;11(24):12069. [doi: 10.3390/app112412069]
32. Valenčič E, Beckett E, Collins CE, Koroušić Seljak B, Bucher T. Digital nudging in online grocery stores: a scoping review on current practices and gaps. Trends Food Sci Technol 2023 Jan;131:151-163. [doi: 10.1016/j.tifs.2022.10.018]
33. Yera R, Alzahrani AA, Martínez L, Rodríguez RM. A systematic review on food recommender systems for diabetic patients. Int J Environ Res Public Health 2023 Feb 27;20(5):4248. [doi: 10.3390/ijerph20054248] [Medline: 36901271]
34. Forberger S, Wichmann F, Comito CN. Nudges used to promote physical activity and to reduce sedentary behaviour in the workplace: results of a scoping review. Prev Med 2022 Feb;155:106922. [doi: 10.1016/j.ypmed.2021.106922] [Medline: 34933021]
35. Uta M, Felfernig A, Le VM, et al. Knowledge-based recommender systems: overview and research directions. Front Big Data 2024;7:1304439. [doi: 10.3389/fdata.2024.1304439] [Medline: 38469430]
36. Sun Y, Zhou J, Ji M, Pei L, Wang Z. Development and evaluation of health recommender systems: systematic scoping review and evidence mapping. J Med Internet Res 2023 Jan 19;25:e38184. [doi: 10.2196/38184] [Medline: 36656630]
37. Forberger S, Reisch LA, van Gorp P, et al. 'Let me recommend… ': use of digital nudges or recommender systems for overweight and obesity prevention-a scoping review protocol. BMJ Open 2024 Jul 31;14(7):e080644. [doi: 10.1136/bmjopen-2023-080644] [Medline: 39089719]
38. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
39. Pollock D, Davies EL, Peters MDJ, et al. Undertaking a scoping review: a practical guide for nursing and midwifery students, clinicians, researchers, and academics. J Adv Nurs 2021 Apr;77(4):2102-2113. [doi: 10.1111/jan.14743] [Medline: 33543511]
40. Tremblay MS, Aubert S, Barnes JD, et al. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75. [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
41. Thivel D, Tremblay A, Genin PM, Panahi S, Rivière D, Duclos M. Physical activity, inactivity, and sedentary behaviors: definitions and implications in occupational health. Front Public Health 2018;6:288. [doi: 10.3389/fpubh.2018.00288] [Medline: 30345266]
42. Collings PJ, Backes A, Aguayo GA, Malisoux L, ORISCAV-LUX study group. Device-measured physical activity and sedentary time in a national sample of Luxembourg residents: the ORISCAV-LUX 2 study. Int J Behav Nutr Phys Act 2022 Dec 29;19(1):161. [doi: 10.1186/s12966-022-01380-3] [Medline: 36581944]
43. Peters MDJ, Godfrey CM, McInerney P. Search strategy. In: The Joanna Briggs Institute Reviewers' Manual 2015: Methodology for JBI Scoping Reviews: Joanna Briggs Institute; 2015:13-14 URL: https://www.researchgate.net/publication/294736492_Methodology_for_JBI_Scoping_Reviews [accessed 2026-03-25]
44. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 guideline statement. J Clin Epidemiol 2016 Jul;75:40-46. [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
45. Calvaresi D, Eggenschwiler S, Calbimonte JP, Manzo G, Schumacher M. A personalized agent-based chatbot for nutritional coaching. Presented at: IEEE/WIC/ACM International Conference on Web Intelligence (WI 2021); Dec 14-17, 2021; Melbourne, Australia p. 682-687. [doi: 10.1145/3486622.3493992]




XSL•FO
RenderX

46. Jannach D, Zanker M, Felfernig A, Friedrich G. Content-based recommendation. In: Jannach D, Zanker M, Felfernig A, Friedrich G, editors. Recommender Systems: An Introduction: Cambridge University Press; 2010:51-80. [doi: 10.1017/CBO9780511763113.005]
47. Sanchez-Bocanegra CL, Sanchez-Laguna F, Sevillano JL. Introduction on health recommender systems. Methods Mol Biol 2015;1246:131-146. [doi: 10.1007/978-1-4939-1985-7_9] [Medline: 25417084]
48. Jannach D, Zanker M, Felfernig A, Friedrich G. Collaborative recommendation. In: Jannach D, Zanker M, Felfernig A, Friedrich G, editors. Recommender Systems: An Introduction: Cambridge University Press; 2010:13-50. [doi: 10.1017/CBO9780511763113.004]
49. De Santis KK, Stiens L, Christianson L, et al. What do we know about recommender systems for obesity prevention? Scoping review of reviews within a project HealthyW8 [Poster] [Article in German]. : Dresden Presented at: Gesundheit – Gemeinsam, Kooperationstagung Der Deutschen Gesellschaft Für Medizinische Informatik, Biometrie Und Epidemiologie (GMDS), Deutschen Gesellschaft Für Sozialmedizin Und Prävention (DGSMP), Deutschen Gesellschaft Für Epidemiologie (DGEpi), Deutschen Gesellschaft Für Medizinische Soziologie (DGMS); Sep 8-13, 2024; Deutsches Hygiene-Museum Dresden, Dresden, Germany p. 8-13 URL: https://www.egms.de/de/meetings/gmds2024/ [accessed 2026-03-25]
50. Abbas F, Najjar N, Wilson D. The bites eclectic: critique-based conversational recommendation for diversity-focused meal planning. Presented at: International Conference on Case-Based Reasoning (ICCBR 2021); Sep 13-16, 2021; Salamanca, Spain p. 1-16. [doi: 10.1007/978-3-030-86957-1_1]
51. Ahn J, Williamson J, Gartrell M, Han R, Lv Q, Mishra S. Supporting healthy grocery shopping via mobile augmented reality. ACM Trans Multimedia Comput Commun Appl 2015 Oct 21;12(1S):1-24. [doi: 10.1145/2808207]
52. Chao WY, Hass Z, Gao Q, Zhou J. Choice-based user interface design of a smart healthy food recommender system for nudging eating behavior of older adult patients with newly diagnosed type II diabetes. Presented at: 6th International Conference, ITAP 2020, Held as Part of the 22nd HCI International Conference, HCII 2020; Jul 19-24, 2020; Copenhagen, Denmark p. 221-234. [doi: 10.1007/978-3-030-50249-2_17]
53. Hales S, Turner-McGrievy GM, Wilcox S, et al. Social networks for improving healthy weight loss behaviors for overweight and obese adults: a randomized clinical trial of the social pounds off digitally (Social POD) mobile app. Int J Med Inform 2016 Oct;94:81-90. [doi: 10.1016/j.ijmedinf.2016.07.003] [Medline: 27573315]
54. He Q, Agu E. On11: an activity recommendation application to mitigate sedentary lifestyle. Presented at: Proceedings of the 2014 Workshop on Physical Analytics (WPA '14); Jun 14, 2014; Bretton Woods, New Hampshire, USA p. 3-8. [doi: 10.1145/2611264.2611268]
55. DiCosola BM, Neff G. Nudging behavior change: using in-group and out-group social comparisons to encourage healthier choices. Presented at: 2022 CHI Conference on Human Factors in Computing Systems (CHI '22); Apr 30 to May 5, 2022. [doi: 10.1145/3491102.3502088]
56. Rabbi M, Aung MH, Zhang M, Choudhury T, Acm D, et al. MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. Presented at: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2015); Sep 7, 2015 to Sep 11, 2025; Osaka, Japan p. 707-718. [doi: 10.1145/2750858.2805840]
57. Sengupta A, Bhattacherjee A, Dutta K, Beckie T. Information technology interventions in cardiac rehabilitation: a theory driven approach. Presented at: International Conference on Information Systems (ICIS 2020); Dec 13, 2020 to Dec 16, 2022 URL: https://aisel.aisnet.org/icis2020/is_health/is_health/18/ [accessed 2026-03-09]
58. Yang L, Hsieh CK, Yang H, et al. Yum-Me: a personalized nutrient-based meal recommender system. ACM Trans Inf Syst 2017 Aug;36(1):7. [doi: 10.1145/3072614] [Medline: 30464375]
59. Berger M, Nüske N, Müller C. Digital nudging in online grocery stores - towards ecologically sustainable nutrition. Presented at: 41st International Conference on Information Systems (ICIS 2020); Dec 13-16, 2020 URL: https://aisel.aisnet.org/icis2020/societal_impact/societal_impact/11/ [accessed 2026-03-09]
60. Cromjongh R, Reenen QV, König L, et al. Group adapted avatar recommendations for exergames. Presented at: 31st ACM Conference on User Modeling, Adaptation and Personalization (UMAP 2023); Jun 26-29, 2023; Limassol, Cyprus p. 283-290. [doi: 10.1145/3563359.3597389]
61. Elsweiler D, Trattner C, Harvey M. Exploiting food choice biases for healthier recipe recommendation. Presented at: 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2017); Aug 7-11, 2017; Tokyo, Japan p. 575-584. [doi: 10.1145/3077136.3080826]
62. Starke A, Asotic E, Trattner C. "Serving each user": supporting different eating goals through a multi-list recommender interface. Presented at: Fifteenth ACM Conference on Recommender Systems (RecSys '21); Sep 27 to Oct 1, 2021. [doi: 10.1145/3460231.3474232]
63. Starke A, El Majjodi A, Trattner C, Brusilovsky P, Gemmis M, Felfernig A, et al. Boosting health? examining the role of nutrition labels and preference elicitation methods in food recommendation. Presented at: 9th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS 2022); Sep 22, 2022; Seattle, Washington, United States p. 67-84.
64. Starke AD, Asotic E, Trattner C, Van Loo EJ. Examining the user evaluation of multi-list recommender interfaces in the context of healthy recipe choices. ACM Trans Recomm Syst 2023;1(4):1-31. [doi: 10.1145/3581930]




XSL•FO
RenderX

65. Starke AD, Brynestad EKK, Hauge S, Løkeland LS. Nudging healthy choices in food search through list re-ranking. Presented at: 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP 2021); Jun 26-29, 2021 p. 293-298. [doi: 10.1145/3450614.3464621]
66. Hauptmann H, Leipold N, Madenach M, et al. Effects and challenges of using a nutrition assistance system: results of a long-term mixed-method study. User Model User-Adap Inter 2022 Nov;32(5):923-975. [doi: 10.1007/s11257-021-09301-y]
67. Leipold N, Madenach M, Schäfer H, Lurz M, Terzimehić N, Groh G, et al. Nutrilize a personalized nutrition recommender system: an enable study. Presented at: 3rd International Workshop on Health Recommender Systems (HealthRecSys 2018); Oct 6, 2018; Vancouver, Canada p. 24-29 URL: https://ceur-ws.org/Vol-2216/healthRecSys18_paper_5.pdf [accessed 2026-03-09]
68. Lin Y, Jessurun J, Vries B, Timmermans H. Motivate: towards context-aware recommendation mobile system for healthy living. Presented at: 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops; May 23-26, 2011; Dublin, Ireland p. 250-253. [doi: 10.4108/icst.pervasivehealth.2011.246030]
69. Michels L, Ochmann J, Schmitt K, Laumer S, Tiefenbeck V. Salience, transparency, and self-nudging: a digital nudge to promote healthier food product choices. Eur J Inf Syst 2024 Sep 2;33(5):717-747. [doi: 10.1080/0960085X.2023.2229787]
70. Willinger L, Oberhoffer-Fritz R, Ewert P, Müller J. Digital Health Nudging to increase physical activity in pediatric patients with congenital heart disease: a randomized controlled trial. Am Heart J 2023 Aug;262:1-9. [doi: 10.1016/j.ahj.2023.04.001] [Medline: 37030491]
71. Teuber M, Leyhr D, Moll J, Sudeck G. Nudging digital physical activity breaks for home studying of university students-a randomized controlled trial during the COVID-19 pandemic with daily activity measures. Front Sports Act Living 2022;4:1024996. [doi: 10.3389/fspor.2022.1024996] [Medline: 36506718]
72. Schäfer H, Willemsen MC. Rasch-based tailored goals for nutrition assistance systems. Presented at: 24th International Conference on Intelligent User Interfaces (IUI 2019); Mar 17-20, 2019. [doi: 10.1145/3301275.3302298]
73. Schomaker J, Vriens M, Jarva HA. Healthy or not: influencing attention to bias food choices. Food Qual Prefer 2022 Mar;96:104384. [doi: 10.1016/j.foodqual.2021.104384]
74. Coppens I, De Pessemier T, Martens L. Connecting physical activity with context and motivation: a user study to define variables to integrate into mobile health recommenders. User Model User-Adap Inter 2024 Mar;34(1):147-181. [doi: 10.1007/s11257-023-09368-9]
75. Coppens I, Martens L, De Pessemier T. Motivating people to move more with personalized activity and tip recommendations: a randomized controlled trial. Presented at: IUI '23 Companion: Companion Proceedings of the 28th International Conference on Intelligent User Interfaces; Mar 27-31, 2023; Sydney, New South Wales, Australia p. 123-126. [doi: 10.1145/3581754.3584149]
76. Coppens I, Martens L, Pessemier TD. Analyzing accuracy versus diversity in a health recommender system for physical activities: a longitudinal user study. Presented at: Proceedings of the 17th ACM Conference on Recommender Systems (RecSys 2023); Sep 18-22, 2023; Singapore, Singapore p. 1146-1151. [doi: 10.1145/3604915.3610650]
77. De Bauw M, De La Revilla LS, Poppe V, Matthys C, Vranken L. Digital nudges to stimulate healthy and pro-environmental food choices in E-groceries. Appetite 2022 May 1;172:105971. [doi: 10.1016/j.appet.2022.105971] [Medline: 35181380]
78. Vanhecke K, All A, Martens L, et al. Recipe recommendations for individual users and groups in a cooking assistance app. Appl Intell 2023;53:27027-27043. [doi: 10.1007/s10489-023-04909-6]
79. Gutiérrez F, Cardoso B, Verbert K, et al. PHARA: a personal health augmented reality assistant to support decision-making at grocery stores. Presented at: 2nd International Workshop on Health Recommender Systems (HealthRecSys'17); Aug 31, 2017; Como, Italy p. 10-13 URL: https://ceur-ws.org/Vol-1953/healthRecSys17_paper_7.pdf [accessed 2026-03-09]
80. Castiglia G, El Majjodi A, Calò F, Deldjoo Y, Narducci F, Starke A, et al. Nudging towards health in a conversational food recommender system using multi-modal interactions and nutrition labels. Presented at: 4th Edition of Knowledge-aware and Conversational Recommender Systems (KaRS) Workshop @ RecSys 2022; Sep 18-23, 2022; Seattle, Washington, USA p. 29-35 URL: https://ceur-ws.org/Vol-3294/long1.pdf [accessed 2026-03-09]
81. Elahi M, Ge M, Ricci F, Fernández-Tobías I, Berkovsky S, David M, et al. Interaction design in a mobile food recommender system. Presented at: 9th ACM Conference on Recommender Systems (RecSys 2015); Sep 16-20, 2015; Vienna, Austria p. 49-52 URL: https://shlomo-berkovsky.github.io/files/pdf/RecSys15c.pdf [accessed 2026-03-09]
82. Massimo D, Elahi M, Ricci F, Ge M. Item contents good, user tags better: empirical evaluation of a food recommender system. Presented at: 25th Conference on User Modeling, Adaptation and Personalization (UMAP 2017); Jul 3-12, 2017; Bratislava, Slovakia p. 373-374. [doi: 10.1145/3079628.3079640]
83. Musto C, Starke AD, Trattner C, Rapp A, Semeraro G. Exploring the effects of natural language justifications in food recommender systems. Presented at: Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21); Jun 21-25, 2021; Utrecht, Netherlands p. 147-157. [doi: 10.1145/3450613.3456827]
84. Musto C, Trattner C, Starke A, Semeraro G. Towards a knowledge-aware food recommender system exploiting holistic user models. Presented at: 28th ACM Conference on User Modeling, Adaptation and Personalization (UMAP 2020); Jul 14-17, 2020; Genoa (Genova), Italy p. 333-337. [doi: 10.1145/3340631.3394880]




XSL•FO
RenderX

85. Afonso L, Rodrigues R, Castro J, et al. A mobile-based tailored recommendation system for parents of children with overweight or obesity: a new tool for health care centers. Eur J Investig Health Psychol Educ 2020 Aug 7;10(3):779-794. [doi: 10.3390/ejihpe10030057] [Medline: 34542511]
86. Martinez-Martin E, Costa A, Cazorla M. PHAROS 2.0-A PHysical Assistant RObot System improved. Sensors (Basel) 2019 Oct 18;19(20):4531. [doi: 10.3390/s19204531] [Medline: 31635278]
87. Ribeiro D, Machado J, Ribeiro J, Vasconcelos MJM, Vieira EF, Barros AC, et al. SousChef: mobile meal recommender system for older adults. Presented at: Proceedings of the 3rd International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2017); Feb 21-23, 2017; Porto, Portugal p. 36-45. [doi: 10.5220/0006281900360045]
88. Samagaio AM, Cardoso HL, Ribeiro D, Univ Porto F, Fraunhofer Portugal PP, Lab Inteligencia A, et al. A chatbot for recipe recommendation and preference modeling. Presented at: 20th EPIA Conference on Artificial Intelligence (EPIA 2021); Sep 7-9, 2021; Virtual event p. 389-402. [doi: 10.1007/978-3-030-86230-5_31]
89. Showafah M, Sihwi SW, - W. Ontology-based daily menu recommendation system for complementary food according to nutritional needs using naïve Bayes and TOPSIS. Int J Adv Comput Sci Appl 2021;12(11):638-645. [doi: 10.14569/IJACSA.2021.0121173]
90. Norouzi S, Kamel Ghalibaf A, Sistani S, et al. A mobile application for managing diabetic patients' nutrition: a food recommender system. Arch Iran Med 2018 Oct 1;21(10):466-472. [Medline: 30415555]
91. Thongyoo P, Anantapanya P, Jamsri P, Chotipant S. A personalized food recommendation chatbot system for diabetes patients. Presented at: 17th International Conference of Cooperative Design, Visualization, and Engineering (CDVE 2020); Oct 25-28, 2020; Bangkok, Thailand p. 19-28. [doi: 10.1007/978-3-030-60816-3_3]
92. Sookrah R, Dhowtal JD, Devi Nagowah S. A DASH diet recommendation system for hypertensive patients using machine learning. Presented at: 7th International Conference on Information and Communication Technology (ICoICT 2019); Jul 24-26, 2019. [doi: 10.1109/ICoICT.2019.8835323]
93. Giabbanelli PJ, Deck P, Andres L, Schiphorst T, Finegood DT. Supporting a participant-centric management of obesity via a self-improving health game. Presented at: 22nd International Conference on Human-Computer Interaction (HCI International 2013) — Healthcare and Safety of the Environment and Transport track; Jul 21-26, 2013; Las Vegas, Nevada, USA p. 189-196. [doi: 10.1007/978-3-642-39173-6_23]
94. Haile C, Kirk A, Cogan N, Janssen X, Gibson AM, MacDonald B. Pilot testing of a nudge-based digital intervention (Welbot) to improve sedentary behaviour and wellbeing in the workplace. Int J Environ Res Public Health 2020 Aug 10;17(16):5763. [doi: 10.3390/ijerph17165763] [Medline: 32784956]
95. Gómez-Del-Río N, González-González CS, Toledo-Delgado PA, Muñoz-Cruz V, García-Peñalvo F. Health promotion for childhood obesity: an approach based on self-tracking of data. Sensors (Basel) 2020 Jul 6;20(13):3778. [doi: 10.3390/s20133778] [Medline: 32640540]
96. Alcaraz-Herrera H, Palomares I, Elsweiler D, Ludwig B, Said A, Schafer H, et al. Evolutionary approach for 'healthy bundle' wellbeing recommendations. 2019 Presented at: Proceedings of the 4th International Workshop on Health Recommender Systems co-located with 13th ACM Conference on Recommender Systems (HealthRecSys'19); Sep 16-20, 2019; Copenhagen, Denmark p. 18-23 URL: https://ceur-ws.org/Vol-2439/5-paginated.pdf [accessed 2026-03-09]
97. Li J, Zhang H, He Z, et al. LifeRec: a mobile app for lifelog recording and ubiquitous recommendation. Presented at: CHIIR 2022 – Conference on Human Information Interaction and Retrieval; Mar 14-18, 2022. [doi: 10.1145/3498366.3505837]
98. Nam Y, Kim Y. Individualized exercise and diet recommendations: an expert system for monitoring physical activity and lifestyle interventions in obesity. J Electr Eng Technol 2015 Nov 1;10(6):2434-2441. [doi: 10.5370/JEET.2015.10.6.2434]
99. Palomares I, Alcaraz-Herrera H, Shen KY. F-EvoRecSys: an extended framework for personalized well-being recommendations guided by fuzzy inference and evolutionary computing. Int J Fuzzy Syst 2022;24(6):2783-2797. [doi: 10.1007/s40815-022-01286-z]
100. Svensson M, Höök K, Cöster R. Evaluating social trails. Presented at: ACM CHI 2003 – Conference on Human Factors in Computing Systems; Apr 5-10, 2003; Ft Lauderdale, Florida, USA p. 844-845. [doi: 10.1145/765891.766026]
101. Abdool H, Pooransingh A, Li Y. Recommend my dish: a multi-sensory food recommender. Presented at: IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM 2015); Aug 24, 2015. [doi: 10.1109/PACRIM.2015.7334841]
102. Khan MA, Rushe E, Smyth B, Coyle D, Elsweiler D, Ludwig B, et al. Personalized, health-aware recipe recommendation: an ensemble topic modeling based approach. Presented at: HealthRecSys 2019 – 4th Workshop on Health Recommender Systems; Sep 20, 2019; Copenhagen, Denmark p. 4-10 URL: https://ceur-ws.org/Vol-2439/3-paginated.pdf [accessed 2026-03-09]
103. Khan MA, Smyth B, Coyle D. Addressing the complexity of personalized, context-aware and health-aware food recommendations: an ensemble topic modelling based approach. J Intell Inf Syst 2021 Oct;57(2):229-269. [doi: 10.1007/s10844-021-00639-8]
104. Pecune F, Callebert L, Marsella S. Designing persuasive food conversational recommender systems with nudging and socially-aware conversational strategies. Front Robot AI 2021;8:733835. [doi: 10.3389/frobt.2021.733835] [Medline: 35127834]




XSL•FO
RenderX

105. Grace K, Finch E, Gulbransen-Diaz N, Henderson H. Q-chef: the impact of surprise-eliciting systems on food-related decision-making. Presented at: CHI 2022 – Conference on Human Factors in Computing Systems; Apr 30 to May 5, 2020. [doi: 10.1145/3491102.3501862]
106. Ozeki S, Kotera M, Ishiguro K, Nishimura T, Higuchi K. Recipe recommendation for balancing ingredient preference and daily nutrients. Presented at: Proceedings of the 1st International Workshop on Multimedia for Cooking, Eating, and related APPlications (CEA 2022), co-located with ACM Multimedia 2022 (MM '22); Oct 10, 2022; Lisbon, Portugal p. 11-19. [doi: 10.1145/3552485.3554941]
107. Hamdollahi Oskouei S, Hashemzadeh M. FoodRecNet: a comprehensively personalized food recommender system using deep neural networks. Knowl Inf Syst 2023;65(9):3753-3775. [doi: 10.1007/s10115-023-01897-4]
108. Reinholdsson T, Hedesström M, Ejelöv E, et al. Nudging green food: the effects of a hedonic cue, menu position, a warm - glow cue, and a descriptive norm. J Consum Behav 2023 May;22(3):557-568. [doi: 10.1002/cb.2129]
109. Vandeputte J, Herold P, Kuslii M, et al. Principles and validations of an artificial intelligence-based recommender system suggesting acceptable food changes. J Nutr 2023 Feb;153(2):598-604. [doi: 10.1016/j.tjnut.2022.12.022] [Medline: 36894251]
110. El Majjodi A, Starke AD, Trattner C. Nudging towards health? Examining the merits of nutrition labels and personalization in a recipe recommender system. Presented at: 30th ACM Conference on User Modeling, Adaptation and Personalization (UMAP 2022); Jul 4-7, 2022; Barcelona, Spain p. 48-56. [doi: 10.1145/3503252.3531312]
111. Garcia MB. Plan-cook-eat: a meal planner app with optimal macronutrient distribution of calories based on personal total daily energy expenditure. Presented at: 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM); Nov 29 to Dec 1, 2019. [doi: 10.1109/HNICEM48295.2019.9073490]
112. Garcia MB, Mangaba JB, Tanchoco CC. Acceptability, usability, and quality of a personalized daily meal plan recommender system: the case of virtual dietitian. Presented at: 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM); Dec 28-30, 2021. [doi: 10.1109/HNICEM54116.2021.9732056]
113. Hafez MM, Redondo RPD, Vilas AF, Pazó HO. Multi-criteria recommendation systems to foster online grocery. Sensors (Basel) 2021 May 28;21(11):3747. [doi: 10.3390/s21113747] [Medline: 34071344]
114. Jesse M, Jannach D, Gula B. Digital nudging for online food choices. Front Psychol 2021;12:729589. [doi: 10.3389/fpsyg.2021.729589] [Medline: 34987443]
115. Herrmanny K, Löppenberg S, Schwarz M, Brusilovsky P, Gemmis M, Felfernig A, et al. Investigating mechanisms for user integration in the activity goal recommendation process by interface design. Presented at: 6th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS 2019); Sep 19, 2019; Copenhagen, Denmark p. 46-54 URL: https://ceur-ws.org/Vol-2450/paper6.pdf [accessed 2026-03-09]
116. Ono S, Yotsuya Y, Takahashi N, Sakamoto T, Kato T. Physical-exercise recommendation system based on individual daily schedule. Presented at: 2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS); Nov 29 to Dec 2, 2022. [doi: 10.1109/SCISISIS55246.2022.10001861]
117. Zhao Z, Arya A, Orji R, Chan G. Effects of a personalized fitness recommender system using gamification and continuous player modeling: system design and long-term validation study. JMIR Serious Games 2020 Nov 17;8(4):e19968. [doi: 10.2196/19968] [Medline: 33200994]
118. Starke AD, Willemsen MC, Trattner C. Nudging healthy choices in food search through visual attractiveness. Front Artif Intell 2021;4:621743. [doi: 10.3389/frai.2021.621743] [Medline: 33969288]
119. El Majjodi A, Starke AD, Trattner C. Integrating digital food nudges and recommender systems: current status and future directions. IEEE Access 2025;13:123002-123017. [doi: 10.1109/ACCESS.2025.3588663]
120. Li R, Zhang Y, Cai X, et al. The nudge strategies for weight loss in adults with obesity and overweight: a systematic review and meta-analysis. Health Policy 2021 Dec;125(12):1527-1535. [doi: 10.1016/j.healthpol.2021.10.010] [Medline: 34772518]
121. Purohit AK, Holzer A, editors. Unhooked by design: scrolling mindfully on social media by automating digital nudges. Presented at: Twenty-Seventh Americas Conference on Information Systems (AMCIS) 2021; Aug 9-13, 2021 URL: https://libra.unine.ch/server/api/core/bitstreams/cd8c2ccf-ca8c-42e9-b3e2-2100acbbe09d/content [accessed 2026-03-09]
122. Felfernig A, Wundara M, Tran TNT, et al. Recommender systems for sustainability: overview and research issues. Front Big Data 2023;6:1284511. [doi: 10.3389/fdata.2023.1284511] [Medline: 37965497]
123. Murphy S, Manzano Ó, Brown KN, editors. Design and evaluation of a constraint-based energy saving and scheduling recommender system. Presented at: 21st International Conference of Principles and Practice of Constraint Programming (CP 2015); Aug 31 to Sep 4, 2015. [doi: 10.1007/978-3-319-23219-5_47]
124. Lohmann PM, Gsottbauer E, Farrington J, Human S, Reisch LA. Choice architecture promotes sustainable choices in online food-delivery apps. PNAS Nexus 2024;3(10):pgae422. [doi: 10.1093/pnasnexus/pgae422] [Medline: 39372540]
125. Meier J, Andor MA, Doebbe FC, Haddaway NR, Reisch LA. Review: do green defaults reduce meat consumption? Food Policy 2022 Jul;110:102298. [doi: 10.1016/j.foodpol.2022.102298]
126. Dewies M, Reisch LA. META BI: a tool for describing behavioural interventions. Behav Public Policy 2025:1-28. [doi: 10.1017/bpp.2025.10015]




XSL•FO
RenderX

127. Reisch LA, Sunstein CR, Andor MA, Doebbe FC, Meier J, Haddaway NR. Mitigating climate change via food consumption and food waste: a systematic map of behavioral interventions. J Clean Prod 2021 Jan;279:123717. [doi: 10.1016/j.jclepro.2020.123717]
128. Sunstein CR. Choice engines and paternalistic AI. Humanit Soc Sci Commun 2024;11:888. [doi: 10.1057/s41599-024-03428-0]
129. Kaiser M, Lohmann PM, Ochieng P, Shi B, Sunstein CR, Reisch LA. Leveraging LLMs for predictive insights in food policy and behavioral interventions. Discov Food 2025;5:262. [doi: 10.1007/s44187-025-00552-x]
130. Atas M, Felfernig A, Polat-Erdeniz S, Popescu A, Tran TNT, Uta M. Towards psychology-aware preference construction in recommender systems: overview and research issues. J Intell Inf Syst 2021 Dec;57(3):467-489. [doi: 10.1007/s10844-021-00674-5]
131. Lubos S, Felfernig A, Tautschnig M. An overview of video recommender systems: state-of-the-art and research issues. Front Big Data 2023;6:1281614. [doi: 10.3389/fdata.2023.1281614] [Medline: 37965498]
132. de Castro BA, Levens SM, Sullivan M, Shaw G. Recommender systems use in weight management mHealth interventions: a scoping review. Obes Rev 2025 Mar;26(3):e13863. [doi: 10.1111/obr.13863] [Medline: 39538385]
133. van der Ploeg HP, Hillsdon M. Is sedentary behaviour just physical inactivity by another name? Int J Behav Nutr Phys Act 2017 Oct 23;14(1):142. [doi: 10.1186/s12966-017-0601-0] [Medline: 29058587]

## Abbreviations

**PA:** physical activity
**PCC:** Population, Concept, Context
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
**SB:** sedentary behavior

*Edited by M Balcarras; submitted 26.Feb.2025; peer-reviewed by AM Moraitis, T Baranowski; revised version received 20.Jan.2026; accepted 20.Jan.2026; published 17.Apr.2026.*

*Please cite as:*
*Forberger S, Reisch LA, Van Gorp P, Stahl C, Christianson L, Halimi J, De Santis KK, Lo C, Omane CA, Loyola-Leyva A, Malisoux L, de-Magistris T, Bohn T*
*Mapping Digital Nudges and Recommender Systems for Obesity Prevention: Scoping Review*
*Interact J Med Res 2026;15:e73151*
*URL: https://www.i-jmr.org/2026/1/e73151*
*doi:10.2196/73151*

© Sarah Forberger, Lucia A Reisch, Pieter Van Gorp, Christoph Stahl, Lara Christianson, Jihan Halimi, Karina Karolina De Santis, Chungwan Lo, Cassandra A Omane, Alejandra Loyola Leyva, Laurent Malisoux, Tiziana de-Magistris, Torsten Bohn. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 17.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Emerging Technologies and Vulnerabilities in Older Adults Without Cognitive Impairments: Systematic Review of Qualitative Evidence

Annachiara Fasoli[1], PhD; Maria De Luca[1], MSc; Giorgia Beretta[1], MSc; Chris Gastmans[2*], PhD; Virginia Sanchini[1,2*], PhD

[1]Department of Oncology and Hemato-Oncology, University of Milan, Via Santa Sofia 9/1, Milan, Italy

[2]Centre for Biomedical Ethics and Law, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

[*]these authors contributed equally

**Corresponding Author:**
Virginia Sanchini, PhD
Department of Oncology and Hemato-Oncology, University of Milan, Via Santa Sofia 9/1, Milan, Italy

## Abstract

**Background:** Aged care has recently undergone major transformations due to demographic aging and the concomitant need to manage health care costs. New emerging technologies (ETs) have started to play central roles in the daily management of older adults. For these transformations to effectively promote successful and active aging, it is essential to understand the opinions of older adults on the impact that technology can have on their vulnerabilities and aging process.

**Objective:** This work aims to study the ethically related impact of ETs on cognitively healthy older adults' vulnerabilities.

**Methods:** Using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, we conducted a systematic review of empirical (qualitative) evidence exploring the relationship between ETs and older adults' vulnerabilities as perceived by older adults (older than 65 years) without cognitive impairments. Five major databases (PubMed, Web of Science, Embase, CINAHL, and Philosopher's Index) were queried on March 1, 2022. After eliminating duplicates, titles, abstracts, and full texts were screened for relevance. Data analysis and synthesis followed the preparatory steps of the coding process detailed in the Qualitative Analysis Guide of Leuven methodology, which involved carefully reading the publications included, identifying significant themes, and constructing conceptual schemes for each paper. The quality of the publications was evaluated by using the Critical Appraisal Skills Program.

**Results:** A total of 11,631 results were obtained. Eventually, 70 articles were included, and of these, 46 articles had a high level of methodological quality. The remaining 24 articles had moderate quality. ETs appeared to have an ambivalent effect, mitigating some already existing vulnerabilities, and at the same time, worsening already existing vulnerabilities or creating new vulnerabilities. For example, unconventional monitoring techniques (eg, wearables) often mitigated relational vulnerability, helping to maintain independence and remain at home and in one's community. Conversely, these same devices may negatively affect moral vulnerability, threatening older adults' privacy linked to data confidentiality.

**Conclusions:** This systematic review, which focused on the perceptions of older adults without cognitive impairments, enriches the vast literature about the everyday management and care of seniors by exploring the ethical implications of ETs. This research is complementary to another systematic review of qualitative evidence, which analyzed the views of older people with cognitive disorders on the same topic. Although a certain ambivalence in the use of ETs was identified by both population groups, it is interesting how cognitively healthy older adults give more importance to some dimensions of vulnerability, such as the moral and relational ones, which, in the case of cognitively impaired older adults, are not as significant. Two important aspects identified were the respect of privacy and data security, and the perceived risk of control and surveillance linked to the use of monitoring technologies.

***(Interact J Med Res 2026;15:e69676)*** doi:10.2196/69676

**KEYWORDS**

aged care; emerging technologies; ethical issues; older adults; vulnerability; impairments; PRISMA

## Introduction

### Background

On December 14, 2020, the United Nations General Assembly declared 2021 - 2023 the decade of healthy aging [1]. Healthy aging, as described by the World Health Organization (WHO), aims to improve older adults' lives globally [2,3] by focusing on not only individuals' functional capacities and residual abilities but also the broader environmental context.



XSL•FO
RenderX

The world population is aging rapidly, with 9% of the global population being older than 64 years [4], and in Europe, 19% of people are older than 65 years [4]. By 2050, these figures will rise to 16% and 28%, respectively. This demographic shift reflects major improvements in public and individual health [5]. However, aging is often associated with increased vulnerability, frailty, psychophysical decline, reduced functional capacity and autonomy, and the onset of chronic conditions or major disabilities. These challenges require health care systems, communities, and individuals to adopt innovative tools and care models to ensure sustainability and long-term manageability [5].

In this context, technology has become a key resource, being increasingly pervasive in society and having a particular significance for aged care. Various terms, for example, AgeTech, ElderTech, SilverTech, and GeronTech [5,6], refer to the health-related industrial sector developing devices that support older adults and formal or informal caregivers. Many emerging technologies (ETs) have entered the market, defined as "radically novel and relatively fast-growing" technologies "characterized by a certain degree of coherence persisting over time and with the potential to exert a considerable impact on the socioeconomic domain(s)" [7]. Some examples considered in this work include wearables (smartwatches, smart bands, and emergency pendants), smart home technologies, assistive robots, and virtual reality [8]. Purely mechanical devices, such as ramps and rails, mobility aids like wheelchairs, and toilet modifications (shower chairs and bath seats), fall outside the scope of this work.

When used in aged care, these technologies are sometimes referred to as "welfare technologies" or "assistive technologies" [9]. Welfare technologies, a subgroup of ETs, can be applied across various social domains (health care, education, work, art, etc) and include technologies intended to enable older adults to remain at home, support independent living, compensate for staff shortages in health care sectors, reduce costs, and enhance self-reliance [9]. Similarly, assistive technologies promote individual functioning and social participation [10]. These technologies support not only healthy aging but also *active aging* [11,12], fostering patient engagement [13], empowerment, and shared decision-making in health care and self-management [14,15]. Such initiatives, by encouraging older adults to actively participate in their own care and preserve their physical, psychological, relational, and emotional capacities, may also reduce health care costs [8] and ease the burden on informal caregivers. Promoting active aging also supports a shift from reactive care to health promotion and disease prevention, aligned with a "salutogenic perspective" [16,17], which emphasizes individuals' holistic reflection on managing the vulnerabilities and stressors of old age and prioritizes well-being [12]. ETs can thus contribute substantially to *ageing-in-place* initiatives [5,18-22] and, for cognitively healthy older adults, to *successful aging* [21], which is understood as maintaining a life within one's own community [18].

As aged care evolves, the home may increasingly become the primary setting for care, delaying institutionalization and helping older adults in maintaining routines consistent with their identity. This transition is supported by ETs and do-it-yourself tools, that is, devices that can be managed autonomously by older adults themselves.

The growing integration of ETs in aged care has led to *gerontechnology*, which studies how advances in technology address older people's needs [23]. Substantial literature has since examined ETs conceptually and philosophically, focusing on their ethical implications in the everyday care of older adults [9,24-28]. Among them, some studies have also explored the perspectives of older adults and carers on the desirable features of ETs [29,30], and numerous quantitative and qualitative studies have analyzed the impact of ETs, especially the ethical challenges they raise [30-34]. Within this body of research, older adult populations are heterogeneous and often categorized based on cognitive functioning. Some studies have focused on individuals with cognitive impairments (eg, mild cognitive impairment, Alzheimer dementia, vascular dementia, Lewy body dementia, and frontotemporal lobar dementia) [35,36], while others have considered cognitively healthy older adults who remain relatively independent [37]. To ensure that transformations in aged care driven by the spread of ETs truly support the well-being and successful aging of all older adults, it is essential to understand how both population groups perceive the impact of ETs on their vulnerabilities and aging more broadly.

## Working Hypothesis

To the best of our knowledge, no study currently provides a systematic overview of the qualitative evidence regarding the perceptions of older adults without cognitive impairments toward ETs in general (across all the different forms, from robots to monitoring technologies), their use, and their ethically related impact, particularly the relationship between ETs and older adults' vulnerabilities. The novelty of this work lies both in its systematic mapping of all available qualitative evidence in an aging population and its attention to the intersection between ETs and older adults' vulnerabilities, a narrow and largely undertheorized area within the broader bioethical debate.

Vulnerability in aged care has been analyzed in multiple facets [38]. Six key types of vulnerabilities have been identified: (1) physical vulnerability (PHV), referring to physiological and pathological bodily decline; (2) psychological vulnerability (PV), encompassing emotional, cognitive, and experiential factors affecting mental health; (3) relational or interpersonal vulnerability (RV), rooted in human ontological interdependence and dependence in real-life contexts; (4) moral vulnerability (MV), which positively fosters dignity, respect, and moral preferences, but may also lead to infantilization, depersonalization, and stigmatization if negatively interpreted; (5) sociocultural, political, and economic vulnerability (SPEV), caused by unfair situational factors; and (6) existential or spiritual vulnerability (ESV), relating to intrinsic existential and spiritual conditions that intensify with aging.

However, a gap remains regarding how ETs influence these vulnerabilities. Using these 6 dimensions of vulnerability as a reference and considering their potential interactions with ETs, this study aims to fill this literature gap by analyzing how ETs may affect older adults' vulnerabilities and by assessing their ethically related impact, understood as contents relating to



XSL•FO
RenderX

ethical issues within well-known ethical approaches (eg, principlism and relational care ethics) [39].

## Methods

### Systematic Review Procedure

We carried out a systematic review of qualitative evidence to gain a deeper understanding of the perceptions and experiences of older adults without cognitive impairments with respect to their vulnerabilities in relation to the use of ETs. Systematic reviews of qualitative evidence are methodologically rigorous reviews presenting an up-to-date, comprehensive overview of qualitative studies, that is, semistructured interviews, focus groups, and informal observations [40].

The analysis followed several steps. First, we formulated the research questions. Second, we conducted a systematic search of the literature, consisting of an electronic database search and a "snowballing" process. Third, we identified relevant publications on the basis of well-defined inclusion and exclusion criteria. Lastly, published research results, which met the inclusion criteria, were analyzed and synthesized, creating conceptual schemes in order to identify and define the ethical arguments in response to our research questions.

### Research Questions

We formulated the following interrelated research questions:

1. What is the ethically related impact of ETs on cognitively healthy older adults' vulnerabilities?
2. What suggestions or strategies do cognitively healthy older adults provide for addressing vulnerabilities related to ETs?

### Literature Search

We developed 3 groups of search terms that guided us in formulating the aforementioned research questions (Multimedia Appendix 1). Group 1 consisted of terms that refer to the type of participant, namely, cognitively healthy older adults. Group 2 consisted of terms that pertain to technology and more specifically to ETs. The terms reported in Groups 1 and 2 were identified on the basis of two previous works [8,41]. We used the expressions "social robots" and "assistive robots" (instead of "social and assistive robots") to intercept as many results as possible. Group 3 consisted of terms that refer to the concept of vulnerability or similar concepts (ie, frailty, fragility, frailness, acceptance, attitude, concerns, discomfort, distress, ethical issues, and ethics). Indeed, as already stated, when associated with the aging population, the concept of vulnerability has a specific connotation, represented in the 6 dimensions presented earlier in the text. However, since there may be some papers that refer, in the content of their reflections, to 1 of the 6 dimensions of vulnerability while nonetheless using a different framing or different expression to refer to it, other terms (eg, distress and discomfort) were added. The keywords "acceptance," "attitude," and "concerns" were used for similar reasons. In qualitative studies involving older adults, ethical issues are often not explicitly presented to participants during interviews, focus groups, or other types of approaches. At the most, these issues may be inferred from the broad questions posed to seniors, who are typically asked simpler questions about their adoption of ETs, their concerns, the perceived benefits and drawbacks, the usability, and similar topics.

The 3 groups of concepts were gathered and expressed according to the properties of each major database (Multimedia Appendix 2). The terms and the search strings were chosen and devised by the first author (AF) in consultation with the other authors (VS and CG). The following 5 electronic databases were queried: PubMed, Web of Science, Embase, CINAHL, and Philosopher's Index. These databases cover the literature in biomedicine, bioethics, philosophy, and (medical) anthropology.

Database queries were conducted on March 1, 2022, and studies not available in English were excluded. Multimedia Appendix 2 presents the number of results returned using the search terms. We used EndNote (version X9; Clarivate Analytics) reference library software to organize the citations of the identified papers, and duplicates were manually deleted.

### Inclusion and Exclusion Criteria

For publications to be included in our systematic review and appraisal, they had to meet a predetermined set of inclusion and exclusion criteria (Multimedia Appendix 3). Screening of the articles was not limited by publication date.

As for the population under study, we included two categories of contributions in our selection of papers: (1) studies using standardized tests to assess cognitive function (eg, Mini Mental State Examination and Montreal Cognitive Assessment) and (2) studies that, while not using formal cognitive assessments, explicitly stated that participants were capable of understanding the study and providing informed consent. In these cases, participants were considered to be not only legally competent (ie, able to consent) but also cognitively competent, as explicitly stated or implied by the authors of selected contributions.

The first (AF) and last (VS) authors screened the titles, abstracts, and full texts of identified papers according to these criteria. The abstract screening was performed separately by the first (AF) and last (VS) authors to make sure that the selection criteria were applied consistently. For 87.3% (3077/3526) of the abstracts, the authors agreed on the items to be included or excluded. For questionable abstracts (449/3526, 12.7%), the first 2 authors (AF and MDL) discussed the candidate abstracts until an agreement was reached.

If the full text of an article was not available, the first author or corresponding author of that article was contacted via email to request a PDF copy. To ensure the search was exhaustive, the "snowball technique" was applied to the reference lists of eligible publications to identify additional potentially relevant publications. The search process was performed according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [42]. The PRISMA checklist is provided in Checklist 1.

All the included publications were from the qualitative research literature; thus, they provided insights on the concept of vulnerability and/or related concepts (ie, frailty, frailness, and fragility) in relation to ETs, as expressed by older adults without cognitive impairments during semistructured interviews or focus groups, or in other settings (eg, informal observations).



XSL•FO
RenderX

The final list of included publications and the descriptions of the characteristics of the included publications are presented in Multimedia Appendices 4 and 5, respectively.

### Quality Appraisal

The quality of the included publications was evaluated by using the Critical Appraisal Skills Program (see the Methodological Quality section). The second author (MDL) assessed each full-text article and constantly discussed the assessments with the first author (AF), who also carefully read and assessed the full texts. When doubts arose, discussions were held until a consensus was reached. This appraisal tool allowed the identification of studies with a high, medium, or low methodological quality. None of the studies were excluded based on methodological quality, because after careful verification, they were all found to be sufficiently adequate.

### Data Extraction and Synthesis

#### *Procedure*

For data extraction and synthesis, we decided to follow the 5 preparatory stages of the coding process detailed in the Qualitative Analysis Guide of Leuven (QUAGOL) [43]. In the first stage, the second author (MDL) thoroughly read and reread all the included publications to familiarize herself with the data and to identify the significant themes described. Ethically related content was identified based on themes and concepts present in a previously completed review of argument-based ethics literature [39] and those stemming from the relational care ethics approach as operationalized in the dignity-enhancing care model of bioethics [44]. In the second stage, the first (AF), second (MDL), and last (VS) authors developed a narrative summary of the highlighted parts. In the third stage, for each publication, a conceptual scheme was created. This scheme summarized the various concepts that are important to answer the research questions. An example of a conceptual scheme is illustrated in Multimedia Appendix 6. Each conceptual scheme was examined separately by the authors (AF, MDL, and VS) to verify that it corresponded to each selected publication. The authors discussed conceptual schemes until they agreed on adequate content. In the fourth stage, these individual conceptual schemes were compared to verify that the content of the conceptual schemes reflected the most important concepts to answer the research questions. All conceptual schemes were transformed into a global scheme that showed the ethics-related impact of ETs on the dimensions of aged care vulnerabilities (Multimedia Appendix 7), as gleaned from the perceptions and experiences of older adults. An iterative analysis and a check between this scheme and the previous stages of QUAGOL were conducted to ensure that the scheme was consistent with the included papers. In the last stage, as reported in this review, we prepared a description of the results.

#### *Categories of ETs*

The ETs used in the qualitative studies under review were organized in this work according to the categories proposed in a previous publication [8]. Conventional monitoring techniques (CMTs) are those technologies that rely on traditional devices, such as oximeters or blood pressure monitors, to continuously observe older adults' physiological and physical parameters and keep track of their condition through technological platforms [8,45-47]. On the other hand, unconventional monitoring techniques (UNMTs) introduce innovative technology components, such as body or environmental sensors, that allow more pervasive monitoring, and they comprise, for instance, smart home technologies or ambient intelligence [48,49]. Virtual reality techniques (VRTs) refer to headsets that create 3D environments in which older adults are "immersed" [50,51]. Finally, socially assistive robots (SARs) are robots that are able to interact with older adults and provide physical or cognitive assistance [52,53].

#### *Conceptual Framework*

In conducting data extraction and synthesis, we adopted a 2-fold strategy (top-down and bottom-up) to assess the ethically related impact of ETs on vulnerabilities. The top-down approach, a deductive and theory-driven method, helped guide and focus our research questions. The approach originally set out was relational care ethics, which emphasizes the moral significance of care relationships and human dependencies. This approach was operationalized in a dignity-enhancing model of care [44] where both dignity and vulnerability played central roles. We also relied on a framework of macro-themes and key issues derived from the relevant literature [8,24,26,39]. This interpretative lens, encompassing several sensitive ethical concerns, such as security, privacy, and autonomy, was applied, refined, and adapted to analyze the selected studies and to assess their ethical impact in relation to older adults' vulnerabilities.

The concept of vulnerability itself was informed by a recent systematic review of argument-based literature [38], which proposed a taxonomy we adopted to present ethically related findings. This allowed for a more analytical understanding of how ETs interact with both existing and potential vulnerabilities in cognitively healthy older adults.

The bottom-up approach relied on an inductive, data-driven method, seeking to infer underlying meaning directly from participants' words and expressions.

## Results

### General Descriptions of Included Publications

A total of 70 publications (Multimedia Appendix 4) met our inclusion criteria: 67 were identified from the research queries and 3 were identified through the snowball technique. For a detailed description of the general characteristics of the included papers, see Multimedia Appendix 5. The PRISMA flowchart is provided in Figure 1.

All the included publications were from the qualitative research literature; thus, they provided insights on the concept of vulnerability and/or related concepts (ie, frailty, frailness, and fragility) in relation to ETs (Multimedia Appendix 3), as expressed by older adults without cognitive impairments during semistructured interviews or focus groups and in other qualitative research settings (eg, informal observations) (Multimedia Appendix 5). With respect to the method used to obtain the qualitative data, 61 (87%) studies used semistructured or general interviews (38/70, 54%) or focus groups (23/70, 33%), and 5 (7%) studies used a combination of semistructured



XSL•FO
RenderX

interviews and focus groups. Other methods included observations (7/70, 10%), usually in combination with interviews or focus groups, and self-report diaries (1/70, 1%).

Publication years ranged from 2004 to 2022, with 53 (76%) articles published between 2015 and 2022. The Americas, Europe, Australia-Oceania, and Asia are represented in this systematic review, meaning that the included studies had first authors from all these different continents. However, most studies were conducted in Europe (36/70, 51%) and the Americas (22/70, 31%). Within Europe and the Americas, the authors' affiliations varied widely. For authors based in Europe, the affiliation for 16 (23%) publications was an institution in Britain or Sweden. For authors based in the Americas, the affiliation for 19 (27%) publications was an institution in the United States.

For inclusion in this review, papers had to involve older adults without cognitive impairments in the context of aged care. All the included papers featured older adults aged 65 years or older. More specifically, 20 (28%) studies included older adults aged 70 years or older, and 2 (3%) included older adults aged 80 years or older. Among the included studies, 47 (67%) were unclear about or did not mention the health status of the participants. Moreover, 26 (37%) studies reported the participants' pathological conditions and/or physical impairments, and 11 (16%) studies indicated cardiovascular pathologies or not-specified chronic conditions.

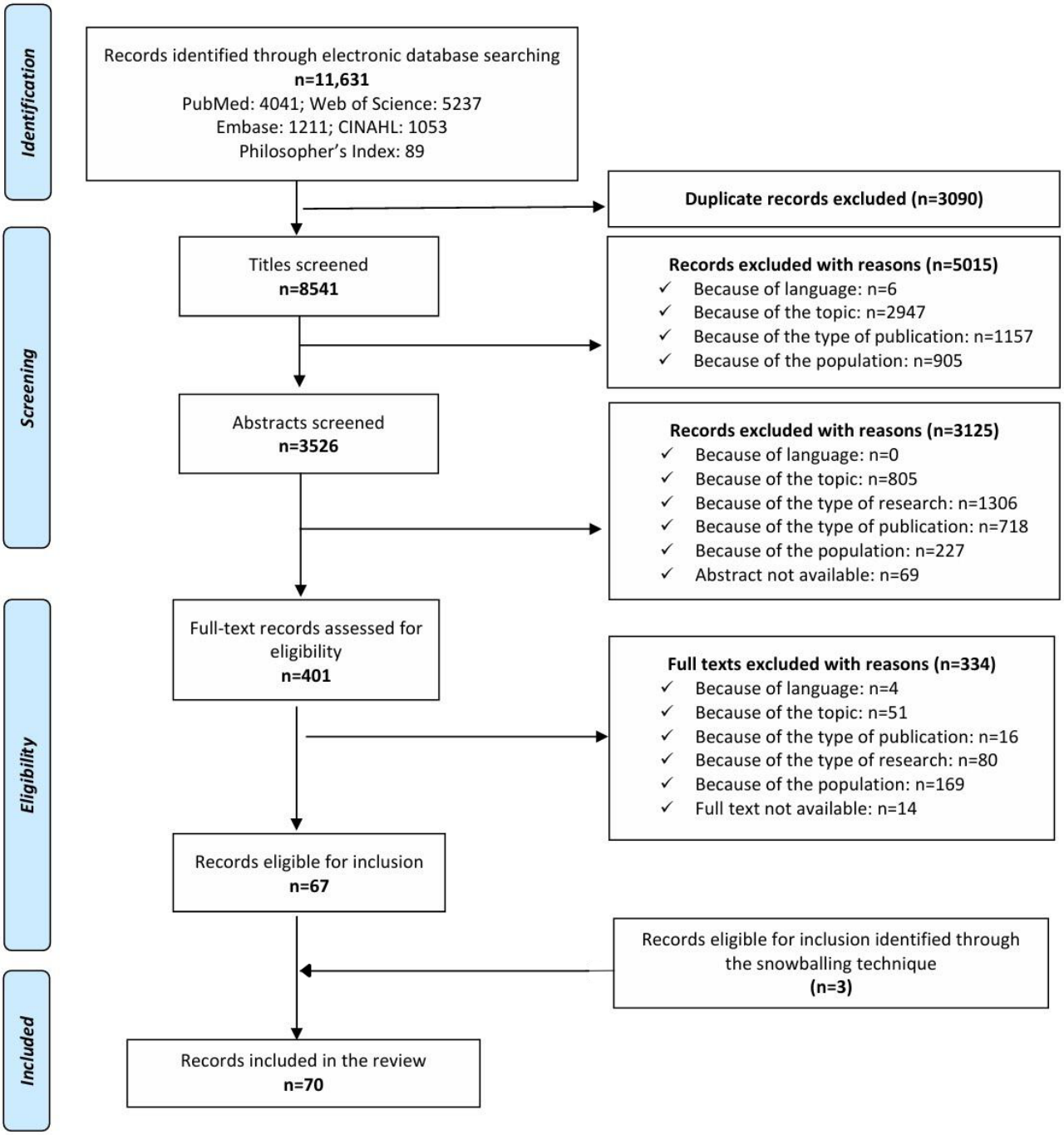
With regard to technologies, all previously mentioned categories of ETs, except for VRTs, were used in the included publications. The most studied category was that of UNMT, which was reported in 41 (58%) publications. Moreover, 29 (41%) studies focused on the use of environmental sensors (eg, chair and bed sensors, motion sensors, and fall detection sensors) [37,54-81]. In the papers, the authors often referred to these monitoring tools using the term "smart home," and they less frequently used "ambient intelligence" or "Internet of Things." In 14 (20%) studies [54-56,67,69-71,78,82-87], the focus was on the use of wearable sensors, such as smartwatches and smart bands, which can collect a variety of information regarding physical activity or vital signs. For example, in 1 study [84], participants used Fitbit, a wearable device that monitors some aspects of the user's behavior and vital parameters (eg, number of steps, distance covered, calories burned, heart rate, and blood pressure).

CMT was the second most studied category, which was noted in 25 (36%) publications. More traditional and common information and communication technologies, such as mobile phones, smartphones, tablets, and computers, were used to provide health-related services (eg, televisits; teleconsultations; and tips for self-management and self-monitoring of one's health, well-being, and/or illness) through the use of so-called mHealth apps [88,89]. The use of such ETs in health care is referred to as "telehealth" [80,90,91].



XSL•FO
RenderX

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart showing the electronic database search, publication identification, and selection process for the included articles.




Finally, the least studied category was SAR, which was noted in 16 (23%) publications. These robots, which were described in detail in 9 (13%) studies [30,92-99], can be divided into 3 groups: nonanthropomorphic [92,93,95,98,99], anthropomorphic [30,95], and zoomorphic [95-97]. An example of the first type of SAR is the robotic platform Hobbit [92,99], which is connected to an ambient assisted living environment and provides entertainment, reminders, fall and emergency detection, and prevention services. Regarding the second group, NAO [95] and Alice [30] are humanoid robots, which are mostly used as companion-type robots to interact and communicate with older adults. Among robots with animal features, the most famous one is PARO [95-97], a baby seal robot that can elicit positive emotions. PARO was originally designed for older adults with dementia, but it has also proven to be a valid tool for older adults without cognitive disorders, as noted in 3 (4%) studies included in this review.

The fourth category of ETs, VRT, was not present in any of the included studies, and the possible reason is that it is still a state-of-the-art technology and has a high cost.

## Methodological Quality

The quality assessment results of all included publications are summarized in Multimedia Appendix 8. According to the evaluation carried out, 46 (66%) studies had a high level of methodological quality. On the other hand, the remaining 24 (34%) studies were classified as having moderate quality.



XSL•FO
RenderX

All the papers had a clear statement of the aims of the research, explained the relevance, clarified the methodology followed, and presented the findings explicitly and clearly. In 66 (94%) studies, the researchers also discussed the contribution of the study to existing knowledge and identified new areas of research.

With regard to data collection and recruitment strategies, 61 (87%) studies took appropriate measures, paying adequate attention to ethical issues, such as obtaining ethical approval from ethics committees and informed consent from participants. On the other hand, with regard to data analysis, 22 (31%) studies did not provide an in-depth description of the analysis process.

Finally, the quality appraisal revealed 2 important limitations. First, in 56 (80%) studies, researchers failed to justify the research design and describe how they decided which method to use. Second, in 64 (91%) studies, researchers did not take into consideration the influence they would have on participants and the research project, thereby presenting a potential risk for bias.

### Ethics-Related Analysis

The included articles were analyzed using a 2-fold strategy (top-down and bottom-up; see the Conceptual Framework section).

Figure 2 provides an illustration to help understand the results of our systematic review in relation to the broader debate surrounding vulnerability in aged care.

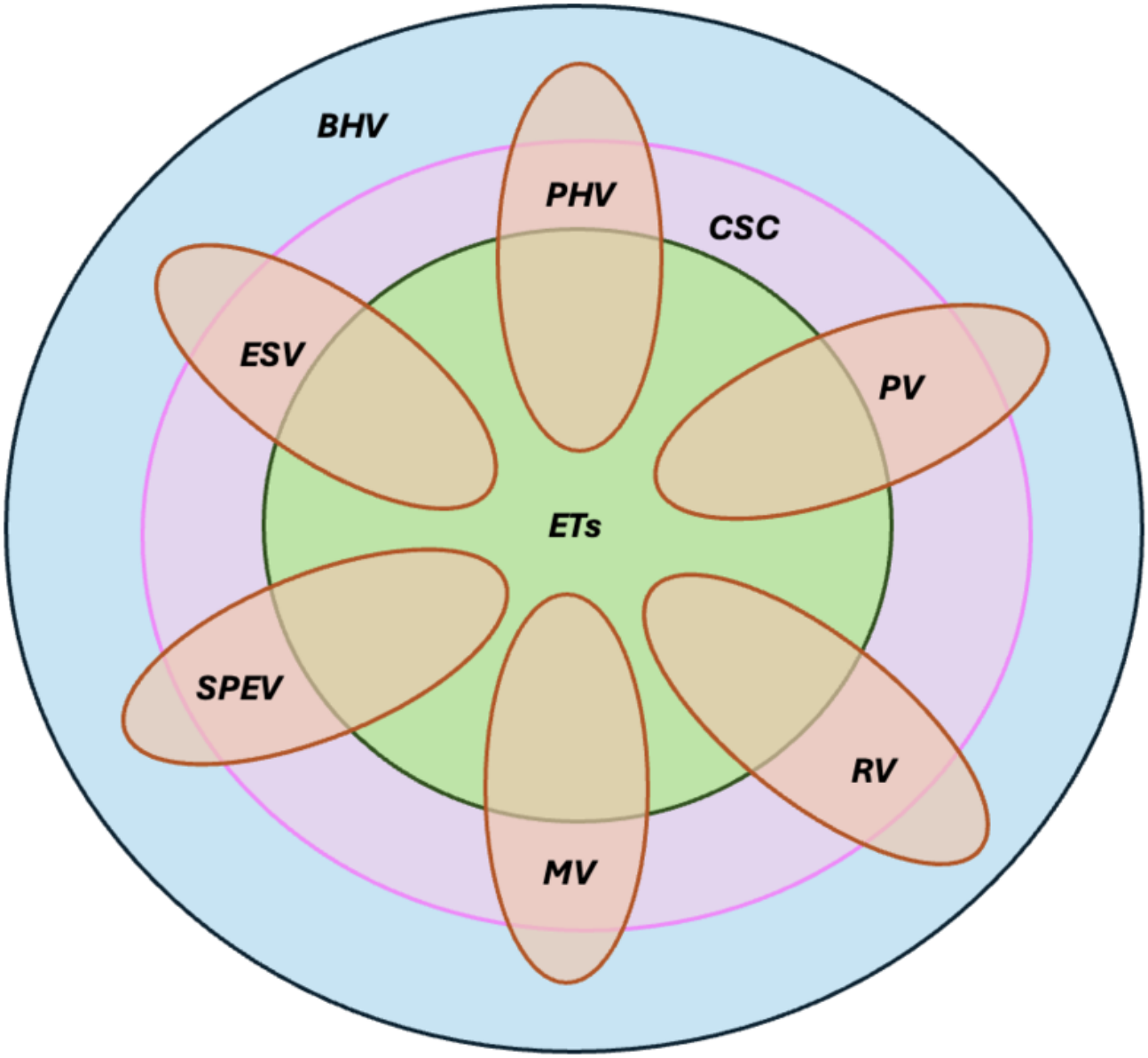
Basic human vulnerability refers to the inherent fragility and finitude of human existence, which cannot be eliminated by technological or other means. In response, humanity has developed various cultural and social constructs, including art, health care, social services, education, politics, economics, and technology, to manage both basic and situational vulnerabilities. These constructs are particularly relevant to the 6 dimensions of vulnerability in aging populations. Among them, ETs may impact both basic and situational vulnerabilities; however, some situational vulnerabilities may resist intervention due to their complexity or deep existential roots.

As summarized in Multimedia Appendix 7, the next sections will present the ethically relevant results of the review, analyzing the impact of ETs on the vulnerabilities of older adults, based on what was reported by study participants. Each of the following paragraphs explores 1 of the 6 dimensions of vulnerability, and they are ordered from the most frequently mentioned dimension to the least frequently mentioned one. Finally, the emerging theme of usability, which was not originally present in the taxonomy, has also been presented.



XSL•FO
RenderX

**Figure 2.** Illustration of the relationship among emerging technologies (ETs), vulnerabilities of older adults, and cultural and social constructs. BHV: basic human vulnerability; CSC: cultural and social constructs; ESV: existential or spiritual vulnerability; MV: moral vulnerability; PHV: physical vulnerability; PV: psychological vulnerability; RV: relational or interpersonal vulnerability; SPEV: sociocultural, political, and economic vulnerability.




## MV Dimension

MV was the dimension of vulnerability most cited by the qualitative studies in this review, with 43 out of the 70 publications addressing it in its different nuances and meanings.

After analyzing the included papers, we further organized this dimension into the following subcategories: *decisional autonomy and/or self-determination* (noted in 14 studies), privacy linked to *confidentiality* (noted in 24 studies), privacy linked to *control and surveillance* (noted in 19 studies), privacy meant as respect for *integrity and identity* (noted in 9 studies), and *stigmatization and infantilization* (noted in 13 studies).

Older people participating in the studies included in this review emphasized the impacts that ETs have on privacy linked to *confidentiality* and more specifically to the management of personal data collected by devices and their disclosure to third parties (ranging from caregivers to health professionals and to potential strangers). This aspect was consistently framed in negative terms by the interviewees. In fact, in all the 24 studies in which the theme of confidentiality emerged, ETs were believed to have a negative impact, as most participants felt that ETs violated their sensitive and personal data. This was especially true for UNMTs, which were studied in 19 out of 24 publications. In particular, two functions of monitoring technologies seemed to be of great concern for older adults: (1) the recording of images and sounds [54,57,58,60,62,64,70] and (2) the collection of personal data [29-58,59-66,67,71,73,74,77,85,86]. For example, in 1 study [54], older adults felt that the use of a camera was intrusive and violating. One participant said, “I wouldn’t want to feel that somebody was watching my activities.” In another study [57], talking about an IP web camera, participants said that it was “too invasive and would make them feel uncomfortable in their own home.” With regard to a smart speaker, a few participants “expressed concerns it would be used to listen into private conversations.” Considering data collection, storage, and transfer, older adults were concerned about data security and



XSL•FO
RenderX

expressed the desire to be protected from possible privacy breaches and from the possible misuse and abuse of their personal data [29,58,59,66,67,71,73,74,77,85,86]. For example, in 1 study [58] that addressed home monitoring techniques, such as wireless motion sensors, in-kitchen temperature sensors, and door contact sensors, a participant stated:

> *I might be concerned about confidentiality issues, such as a breach of my personal information or potential harms due to my information being exposed to others.…There could be a possibility that a burglar could see my data accidentally and come along after I go out.*

In another study [29], in relation to the use of different kinds of technologies based on artificial intelligence, the following statement was made:

> *If they're [the technologies] so sensitive, they know three weeks before we know what's going wrong with our bodies. It seems to me that that kind of information could really be compromised, and seniors could, uh, who are very vulnerable, could really be hoodwinked more easily.*

Participants were also concerned about the transfer of data, which they would like to keep confidential and share only with their relatives and/or caregivers [59,63]. In 1 study [59], a participant said:

> *When you get older, you don't feel that safe on your feet…and you mean – is it – I have to announce that to everybody?…I think that's the most private thing that you're really interfering with.*

In another study [63], older adults had concerns even "about family members potentially 'knowing too much' and some were concerned that data could be used by families to the detriment of the older adults."

The second most recurrent topic concerning MV was privacy linked to *control and surveillance*, which was noted in 19 studies. Among these studies, 16 reported that ETs had a mostly negative impact, 2 [77,83] reported that devices had an ambivalent effect on MV, and 1 [100] mentioned that ETs appeared to mitigate MV.

Among the critical issues identified, the most prominent concern associated with ETs in general (CMTs [101] and SARs [30,102]), but especially UNMTs [54,55,59,62,65,70,73,78,79,84]) is surveillance. For example, in 1 study [55], a participant associated this negative feeling of constant monitoring with past traumatizing experiences, when "there were 'folkpolice' everywhere watching everyone, and everyone watched everyone else in East Germany… and we don't want that system here. And, at least not here in our home." In another study [59], a participant said that "you'd feel like a puppet on a string…I don't want to be – something watching me." Finally, in another study [79], participants linked surveillance to the Big Brother Syndrome [103], and the following statement was made:

> *While they do not mind having the system monitoring them 24/7, the incorporation of a camera into such system would be "too intrusive." "No, definitely no photos. It's like big brother, and i'm not having that."*

In 2 studies, UNMTs were reported to have both positive and negative effects. In 1 study [83], which examined a wrist-worn monitoring device, the positive aspect highlighted was the increased sense of safety, resulting from the awareness of being constantly monitored. For example, a participant said:

> *I feel safe. It must be safer.…In this way they can keep a track of me even when they are not here, which is good because I'm alone here in my house.*

This perception of protection, linked to surveillance, was particularly emphasized in relation to falls, as also stated in another study [100]. On the other hand, in these studies, a strong discomfort also emerged, resulting from the feeling of being "watched," which was perceived as a violation of personal and private space.

Among the included studies, 14 reported that older adults felt that ETs had either a positive or negative impact on their decisional autonomy and/or self-determination. In a positive way, CMTs and UNMTs improved older adults' decisional autonomy and minimized their loss of control, as they enabled older adults to control certain elements of their house or their life [61,76,104-106]. For example, in some studies [61,76], smart home technologies were really appreciated because these devices provided the possibility to remotely control the environment (eg, lights, television, and doors). With regard to GatorTech Smart Home [61], an older woman liked above all "the ability to command the house to open blinds, doors, lights" and "the ability to see who was at the front door and make the choice to see if wanted to respond rather than having to go over and look through the peephole."

However, despite this benefit, ETs (CMTs, UNMTs, and SARs) can also interfere in a negative way with older adults' decisional autonomy, as they direct the life of older adults and make them lose control of their daily activities and decision-making capacity [29,30,55,59,77,94,101,107]. This fear emerges not only in relation to monitoring technologies [29,55,59,77,101,107] but also in relation to the use of SARs [30,94]. Talking about Alice [30], a robot with a small humanoid body and a human-like face, a study mentioned about "negative feelings related to feeling powerless in stopping the perceived technological imperative. Participants felt as if they had no choice but to comply with the evolution of SAR." One participant said that SARs "cannot take over," and another claimed that "not everything can just happen because it [the SAR] says it."

Other frequently mentioned themes in relation to MV were stigmatization and infantilization, which appeared in 13 publications. Twelve publications highlighted how ETs exacerbate feelings of stigmatization and infantilization because older adults associate the use of these devices (mostly UNMTs) with aging and dependency [30,47,55,62,64,74,76,79,80,94,95,104]. For instance, both CMTs and UNMTs [47,62,64,74,76,79,80,104], especially smart home technologies, were despised because the participants believed that these are tools for more vulnerable people with serious disability problems and not devices for healthy



XSL•FO
RenderX

individuals like them. Two studies specifically pointed out the phenomenon of depersonalization, as monitoring techniques like sensors [55] and SARs [30] gave older adults the idea of being seen as an *object* rather than a person with individual needs. With respect to the infantilization phenomenon, the issue emerged negatively in relation to the use of SARs. In particular, in 1 study [94], it was stated:

> *Some participants found the concept of interaction with companion robots demeaning.…Craig was similarly condemnatory of ElliQ: "Crazy. I do think I'm a bit more intelligent than that."*

Likewise, in some studies [30,95], the use of robots was described as childish and dehumanizing.

Only 1 publication showed that ETs could have a positive impact on MV, as, according to the participants of the study, Interaktor, an interactive app for regular reporting of health concerns, allowed health care personnel to take care of older adults as people, reducing depersonalization [46]. Older adults reported a "feeling of being appreciated" and a "feeling of being acknowledged as a valued person."

Shifting the perspective on privacy again, the last 2 connected themes were integrity and identity. In 9 studies where these themes were mentioned, the negative impact of ETs was emphasized, mostly with reference to UNMTs [54,55,57,59,62,71,77], such as smartwatches, health trackers, chair and bed sensors, motion sensors, and home energy sensors. In 8 publications [54,55,57,59,62,71,94,101], the problem was the perceptions of obtrusion, intrusion, and invasion, which were interpreted as violations of personal space, a kind of "attack," compromising personal integrity and identity. In addition, in 1 study [60], a participant was very concerned about the potential misuse of her identity, which was caused by the use of "welfare technology" (ie, ambient assisted living).

### RV Dimension

The second most frequently cited dimension of vulnerability was RV, which appeared in 40 out of the 70 publications. These 40 studies focused on interdependence in real-world settings, and none of them reported an impact on the ontological interdependence characterizing human nature, that is, the intrinsic relational and social nature of human beings.

The relationships among older adults, caregivers, personnel, and/or family members appeared in 36 publications. In 19 studies, older adults highlighted only the positive impacts of ETs on RV. In 7 studies, on the other hand, some disadvantages concerning RV emerged. Finally, in 10 studies, participants emphasized the ambivalent effects of ETs on RV.

Most publications showed that ETs can positively influence RV as participants saw ETs as a support to maintain their independence and remain at home and in their own community [57,65,75,81,94,108,109]. This was especially true for UNMTs like smart home technologies. For example, in 1 study [75], with regard to a sensor monitoring system, which included passive infrared motion sensors, magnetic contact sensors on doors and cabinets, and a flush sensor in the toilet, a participant said:

> *It may be useful for the future.…People can stay at home longer with the help of sensors because there is more supervision.…It is always nice to stay in your own neighborhood, especially for elderly people who have neighbors and friends in their neighborhood.*

Another important benefit with regard to RV was the relief from loneliness [61,65,92,94,106,108,110,111]. SARs, in particular, were seen by older adults as pleasant companions to spend time with. In 1 study [94], ElliQ, a nonmobile home assistant robot with moving parts, conversation functions, and a female-sounding voice, was praised by respondents, and it was stated:

> *…would be comforting in a quiet and lonely household: "It breaks the silence of the day." For Sarah, the more linguistically sophisticated presence of ElliQ had a positive side, because it was like having a person in the house.*

CMTs, such as tablets, smartphones, and laptops, which enable remote communication and video conferencing, were also highly appreciated as ways to connect with members of one's network and receive practical and emotional support from them [37,89,91,98,106,107,112]. In 1 study [89], a participant's point of view on the use of mHealth apps was reported:

> *You'll [the doctor] assist me in keeping an eye on it [the remedy] and reporting back through the app and then I can connect with a trusted advisor just through the app – that would be great. Convenient.*

In another study [112], an interviewee made the following statement about tablets (such as iPad and Kindle Fire): "I feel more informed; I feel I'm in more contact with my family."

In other publications, however, ETs were reported to have a negative impact on the relationships among older adults, caregivers, personnel, and/or family members. According to 1 study [73], UNMTs like ambient assisted living might reduce the independence of older adults, and a female participant stated:

> *I think it's very intrusive and I think people who are over 70 as I am would find it just one step down from you making dependent on somebody…If you feel you've got your independence…you know…I think it would be own step down…yes I didn't like it.*

Sometimes ETs were despised because they can replace human interaction and reduce "face-to-face care" [80,90,100,101], increasing the sense of loneliness [101,113]. Interestingly, ETs were considered problematic because they might generate misunderstandings and mistrust among people [113,114]. For example, in 1 study [113], with regard to the use of a tablet-based communication app allowing asynchronous multimedia communication, it was stated:

> *Six participants verbalized that the technology created family tensions.…Our older people preferred asynchronous communication, to send audio messages, and to receive text messages, but family preferred synchronous communication and video and photo messages. Participants were "disappointed" with relatives that instead of replying to their messages via the app called them on the telephone.*



XSL•FO
RenderX

Finally, as anticipated, in 10 studies with regard to RV, older adults expressed their opinions on ETs in ambivalent terms. ETs appeared to enable older adults to maintain and improve their autonomy and independence in daily activities [30,67,72,77,104,105], but the use of these same tools raised their fear of losing residual autonomy and abilities [76,95,104], that is, the set of psychophysical capacities and functions that are not (for the time being) affected by the deterioration that normally characterizes human aging. Moreover, while ETs appeared to improve contact and provide support [30,46,66], which can reduce feelings of loneliness [30,76,95], they may exacerbate this same loneliness [105] and raise fears of a gradual loss of human interaction [30,46,66,67,72,77].

With regard to human-robot interaction, in 4 studies [93,97,108,115], older adults only emphasized the positive aspects of using ETs. In 4 other studies [71,91,92,100], however, only critical issues emerged. In 3 studies [30,94,95], older adults' opinions were ambivalent. Paradoxically, the use of SARs was appreciated for contradictory reasons. On the one hand, participants appreciated the resemblance of robots to living beings (human beings [95] or animals [94]). For example, in 1 study [94], talking about Biscuit, an animal-like robot, an interviewee said that it showed compassion and "a dog-like affection," and the interviewee made the following comment:

> *You could touch Biscuit.…It's got eyes you can look in. So, you get some sort of empathy back, or feeling back, the way it tilts his head when you talk to him.…More like a real dog.*

In 1 study [95], with regard to home robotic devices, it was mentioned that "participants reacted to the nonverbal communication capabilities of some robots, stating: It's like having another person in the flat that would communicate with me." However, other older adults held the opposite view, and it was stated that robots are more appreciable because they do not have the "defects" of living beings, both human [30,108] and animal [97]. For instance, in 1 study [108], the following statement was made: "the fact that the robot could not 'speak, offend, or get angry like a human being' was mentioned as an advantage of this special companion by one participant." Likewise, in another study [97], PARO, a baby seal robot, was appreciated because, as a participant stated, "you don't have to feed him or clean his litter box or sweep up his hair from rugs." In another study [30], the humanoid robot Alice was described as follows:

> *Very accurate. They will not be hasty, or they will not have a bad mood.…They won't forget anything.…They will not be distracted, will not have a headache. I think it is positive that they will always be 100%.…Available.…Concentrated.…Trustworthy.*

In addition, other emphasized aspects were features like a nice physical appearance [93,97], inviting and "unintimidating" attitudes and cuteness [97], and relational support [115].

These results, however, were mirrored by criticism. While some older adults appreciated SARs for their "roboticism" and "nonsimilarity" to living beings, others despised the fact that they are programmed to "pretend" to be what they are not. Indeed, some older adults refused their inauthenticity and lack of spontaneity [30,92,94,95,100]. Their simulation of living beings' capabilities (affective, social, etc) was strongly rejected as phony. For example, in 1 study [94], ElliQ, a nonmobile robot, and Vector, a toy-like robot, were described as follows:

> *…lacking feeling and being "cold."…Stephanie said: …I don't' like the sort of seeming humanizing of the whole thing. It's ridiculous.*

Even Biscuit [94], which was praised for its "dog-like affection" by 1 older adult, was criticized by another:

> *…the simulation of a living dog and the reactions it prompts are phony.…A live dog is "not programmed to be joyous when it sees you" (Sarah).…Gwen…found Biscuit repellent because, "It's false. It's something that I'm trying to be happy or I'm trying to be something that I'm not."*

This criticism was reiterated in another study [95], where an older adult said that "the only thing that a robot could do to me or respond to me would be a response that's been built in, it's not spontaneous." The critique of the lack of authenticity was connected to the rejection of the robot-provided mode of care, as stated in a study [30], in response to the use of Alice:

> *…only could interact in a preprogrammed manner.…According to most participants the care SARs provide to older adults would lack a fundamental quality of human care: "empathy," "care for the hearth" and so forth.*

In 2 papers [71,91], the issue of human-robot interaction emerged in connection with the use of CMTs and UNMTs. In particular, older people complained about a sense of anxiety and doubt. In the first study [71], talking about some digital services, such as Amazon and Google, activity trackers like Fitbit, and home energy sensors like Nest, a participant said:

> *Although I use the computer, I do find it quite frightening. The reason is that I don't understand it. And I don't know how to put things right.*

In the second study [91], it was mentioned that telehealth technologies, such as videoconferencing, elicited anxiety for various reasons: "(1) receiving spam, (2) experiencing system updates, (3) losing written text, (4) damaging a device, (5) fearing the use of technology in general, (6) fearing microwave radiation, (7) fearing inadequate privacy protection, (8) feeling unsafe using the internet, and (9) fearing online scams or cyber criminals."

### PHV Dimension

PHV was reported in 36 out of 70 publications, thus constituting the third most frequently cited dimension by participants in qualitative studies. An analysis of these 36 studies showed that in most cases, ETs appeared to have a positive impact on PHV.

Indeed, 33 studies revealed that ETs mitigated PHV, which was interpreted as nonpathological physical or physiological bodily deterioration related to aging. Participants appreciated ETs, particularly UNMTs, for different reasons. ETs with monitoring functions can detect emergencies like accidents and falls [30,37,46,58,62,64,66,75,79,81,87,99,100,116] and can monitor health conditions, daily habits, therapy compliance, and health



XSL•FO
RenderX

changes [54,56,57,70,75,77,81,83,85,99,101,104]. They were also useful in some cases to monitor the environment and the opening of doors and windows [61,65,70,74,92,100]. These factors, according to older adults, increased the sense of safety and security [37,46,55,65,66,70,72,74,75,77,79,81,83,85,92,100,101,104]. This positive impact was associated with the possibility of remaining in one's own home and avoiding institutionalization in a nursing home [66,77,83,101]. For example, in 1 study [75], the importance of environmental sensors for both fall detection and detection of physical decline was emphasized:

> *Mrs. D explained this as follows: "Well, you are on your own, so something can happen, like when you fall and can't get up.…I had the idea that this should be watched by someone somewhere."…Mrs. A expressed this as follows: "if there should be a slow change in my daily pattern, I certainly wouldn't report it. I wouldn't notice, and therefore, I find it important that the nurse's station gets a signal like: keep an eye on that"*

Similarly, in 1 study [100], the robot Jibo was admired for the home security it provides:

> *P12 told a story of how the robot could help them…"It seems like a useful idea that if I'm sleeping, and I could have a robot that detects something unusual that I would like to be alerted to…You could program the thing and say if I have a sound like somebody trying to get through the window, please wake me up."*

Other participants highlighted how the use of ETs, above all wearable activity trackers, such as Garmin Vivofit 2, Fitbit, Apple Watch, Jawbone, Misfit Nike, and Gear Fit, which automatically track and monitor various indicators of physical activity (eg, steps taken, pulse or heart rate, calories consumed, etc), could motivate them to be more physically active and improve their health status [30,57,75,82,84,86]. For instance, in 1 study [86], it was stated:

> *Participants agreed that tracker use motivated them to walk more, driven by quantifying activity (counting steps) and continuously making users "more conscious of extra walking" (female participant, non-user).*

Finally, the last important positive aspect concerning PHV in terms of nonpathological deterioration was the physical support and assistance SARs can provide [99,115], assisting older adults in house chores and picking up and transporting objects.

Some studies [30,58,61,77,81,88,117] showed that ETs can also mitigate PHV linked to pathological physical age-related conditions. Some participants perceived ETs, mostly UNMTs, as useful tools to prevent diseases and manage illnesses, such as diabetes, heart failure, dementia, and Alzheimer disease [58,77,81,88,117], and physical impairments, such as vision deficits [30,61]. In 2 studies [79,91], however, participants reported that ETs could exacerbate PHV. They were concerned about the possibility of CMTs and UNMTs negatively impacting a person's health, causing diseases (eg, cancer). For example, in 1 study [79], a participant asked:

> *Could the sensor radio waves give you cancer? I think this is what I would be worried about.*

In another study [91], the use of telehealth technologies, such as videoconferencing, elicited similar comments, and older adults expressed their fear of microwave radiation.

## PV Dimension

PV was the fourth most frequently mentioned dimension by older adults, appearing in 33 out of 70 studies. These 33 papers dealt only with strictly psychological dimensions (cognitive, emotional, or both).

The emotional aspects were those most often mentioned by older people, and they appeared in 23 out of 33 publications. Of these studies, 21 showed how ETs could mitigate PV. In only 8 studies, older adults highlighted the negative aspects.

Focusing on the positive comments, according to participants, the use of ETs could have a positive influence on older adults' emotions and improve their mood: CMTs and UNMTs could provide a sense of safety, security, and reassurance [37,61-64,69,74,80,101,107]. With their monitoring functions, these devices not only contribute to improving PHV but also provide peace of mind. For example, in 1 study [62], a participant commented on the use of smart home technologies, saying that "his friend has decided to use such a device to have 'peace of mind.'" ETs in general, but especially SARs, are stimulating tools that older people enjoy using [46,82,86,92-94,97,112]. Hobbit [92], a robotic platform with personal and social functionalities, raised the following comments:

> *He's interesting and amusing.*
>
> *It is great fun when he is working well.*
>
> *When he moves and his head moves, I get happy and compassionate.…He's charming.*

Another SAR, PARO [69,110], was really appreciated because of its calming effect and its positive influence on emotions. The use of other technologies [46,86,106,112] conveyed a sense of achievement and appreciation. Finally, the appearance of the technology, especially in the case of robots that look like living creatures, aroused positive feelings in users (eg, empathy, compassion, and tenderness) [93-95].

On the other hand, 8 studies revealed how ETs, mostly UNMTs, could have a negative effect on PV, generating adverse emotions and negative attitudes, such as irritation, frustration and annoyance [68,94,95,106], fear [55,56,107], reduced sense of safety [37], and self-doubt and lack of self-worth [106]. For instance, the robot ElliQ was criticized for its cold manners and its disturbing intrusiveness. Participants said:

> *It talks all the time.*
>
> *I don't know whether that would drive me mental if it kept interrupting me and telling me what to do…I might want to get an ax and cut it up.*

In 1 study [107], older adults, reflecting on the use of UNMTs and CMTs, stated that they evoked the fear of "not being able



XSL•FO
RenderX

to manage the technology, making a fool of oneself, and having to ask for help." Moreover, the following statement was made: "Also, fear of being forced into something unwanted and of ruining something was described." Finally, another study [106] stated that even the use of more common technologies, such as tablets, smartphones, and laptops, has challenged older adults, who feel incompetent when they are unable to use the devices, experiencing a sense of inadequacy:

> *Jennifer (female, 71, tablet) demonstrate how her frustration with online activity leads to self-doubt: …"Is it me doing something wrong?"*

With regard to the cognitive level, 8 studies [30,57,67,82,99,102,108,118] showed how ETs could have a positive effect, compensating for memory deficits through the reminder function [30,57,67,99,102,108], influencing motivation and awareness [82], or improving intellectual functions [118]. For example, in 1 study [67] analyzing CMTs and UNMTs like smartwatches, an older adult said:

> *I think it's good because there's some people who as time goes by lose certain of their faculties as time goes by, and memory beginning to fade and so on.*

In another study [82], the Jawbone bracelet, an UNMT that monitors daily activity, was described as follows:

> *I was motivated by the technology, that I freely admit.…I have walked a little more while being monitored.*

In another study [118], participants said:

> *It [the tablet] keeps the brain active.*
>
> *Learning any new skill, surely is helping the cognitive function.*

At the same time, paradoxically, 5 studies [56,76,95,113,118] revealed that ETs could have a negative impact on the cognitive aspects of PV, because they could make older adults more aware of their cognitive decline [56,113] or they could cause a loss of intellectual functioning [76,95,118].

### SPEV Dimension

SPEV appeared in 25 publications. According to 1 study [38], sociocultural vulnerability (SCV) arises when older adults experience social exclusion, isolation, marginalization, or stigma, exacerbated by gender discrimination, educational disparities, and low social support. Economic vulnerability (EV) and political vulnerability (POV) include discrimination in health care access and exposure to unjust judicial systems. These 3 SPEV subdimensions will be analyzed separately.

SCV was identified in 13 studies. Most of these focused on the isolation of older adults and their exclusion from social life, highlighting the positive [58,80,104,106,109,111,112,115] or negative [72,77,101,118] impact of ETs (CMTs and UNMTs for the most part) on SCV.

In 8 studies addressing the advantages of ETs, older adults said that these devices can help them integrate into society and reduce the sense of isolation. For example, in 1 study [115], the SAR was appreciated because it could provide company for older adults, in addition to performing house chores: "Robots will be welcome if I can return to my home, where they can bring me water and live and play with me… (Male, participant 13)." CMTs, such as tablets (eg, iPad) [112], were seen as tools to stay connected to not only family and friends, mitigating RV, but also society and the outside world at large:

> *Harold stated: "I feel more informed; I feel I'm in more contact with my family."…Connie said, "I feel like I'm connected to the world."…"Well, I feel like I've come up in the world."…Carol, 80-years-old, said she felt more "modern…I'm up par with other people." Mary, who is 89-years-old, mentioned that she felt like she was "coming into the 21st century."*

However, some studies indicated that ETs may exacerbate SCV by increasing isolation and social exclusion. Specifically, older adults expressed concerns about losing human contact, abandonment, and isolation linked to the growing use of monitoring technologies [72,77]. In 1 study [77], participants said that ambient assisted living technologies "should not replace human care…'Nothing can replace human contact. The more helpless you are, the more you need for people to come and check on you once in a while.'" In another study [101], the use of CMTs (ie, assistive technology services) made older adults fear being "neglected by the official health care system and society." In another study [118], the concern was different, as the cause of isolation was not the excessive use of monitoring technology, but rather older adults' lack of access to it or inability to use it, which prevented them from keeping in touch with others and the society as a whole:

> *"I'd really like to be in the modern world and to be able to manage these things and to be able to access more. I just feel very limited in what I'm doing.…" (G2, P6). Participants in the no computer experience group also stated: G3, P6: "I think we're missing out a lot, because all the information is at hand and we don't know how to collect that information."…G3, P5: "I just think we're like a forgotten generation, that's what I feel like. You want to go in and you want to be able to talk with your family and your grandchildren and not look vacant when they say, I'm going to do this."*

In 2 studies, older people focused on the barriers to ET use. Participants in the first study [104] expressed concerns about access to CMTs and about their inexperience with these devices. In the second study [114], older adults stressed how different understanding and use of technology between young and older people expose intergenerational differences, and the paper stated:

> *Grandchildren preferred to send videos and receive video/images, while participants mostly sent audio and preferred to receive text messages. Additionally, grandchildren would not receive replies within a day and thought it was frustrating. Grandparents showed surprise and confusion when we told them that grandchildren were expecting quicker replies, as they wanted to "take time and think about what to say back."*



XSL•FO
RenderX

A study on home monitoring techniques (wireless motion sensors, in-kitchen temperature sensors, door contact sensors, and pressure mats) considered SCV by examining older adults' living situations [58]. It highlighted the positive contribution of ETs in reducing abandonment [58], as these technologies compensate for the lack of support from children.

EV was addressed in 13 studies. All of them, except for 1 study [89], highlighted how ETs exacerbate this vulnerability. Most older adults' comments indicated that ETs have a strong financial impact on their income. Many participants expressed concerns about the high costs of these devices and their maintenance, which they could not afford [30,57,62,65,66,72,79,83,88,90,105,118]. This criticality emerged for all 3 types of ETs, but especially in relation to UNMTs, such as smart home technologies. Only 1 study reported how mHealth apps mitigate EV, as they could allow older adults to save money on health management [89]:

> *I mean one of the things I think is terrific that we can email our doctors and get an answer quickly. I'm all for it. It makes good sense. Participant 11.*

Finally, POV was discussed in only 1 study [111], and older adults found that the use of CMTs had a negative impact on this vulnerability. The study showed that ETs exacerbate POV because the lack of skills and knowledge on ETs could complicate older adults' participation in political life and decrease their possibility to be informed and obtain economic benefits [111].

### ESV Dimension

Lastly, ESV was by far the least cited, as it appeared in only 1 publication [113]. In this study [113], according to the participants, a CMT, such as a tablet-based communication app, which allows asynchronous multimedia communication, can exacerbate ESV because it makes older adults aware of their existential limitations and of their finitude:

> *They reported how "inadequate" and "limited" they felt. For Ike (aged 74), the technology made his "Parkinson's battles" more noticeable, from eyesight problems to "losing cognitive abilities" when he forgot about some app's "features." The technology emphasized his health status and a compromised sense of personhood and identity: "I was not like this before," he told us.*

### Usability

Beyond the ethical impact of ETs on older adults' vulnerabilities examined through a top-down approach, adopting a bottom-up perspective revealed significant findings regarding usability. Usability issues were addressed in 36 of the 70 publications reviewed. According to the International Organization for Standardization (ISO), usability is "the extent to which a system, product, or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" (ISO 9241-11:2018) [119]. Usability was reported to significantly shape users' perceptions of a technology's usefulness and acceptability [120].

Regarding this topic, we observed that participants' perceptions related to the usability of ETs could be both positive and negative. A total of 18 publications highlighted the advantages of all types of ETs (CMTs, UNMTs, and SARs). However, critical issues were emphasized more often in 25 publications.

The first usability issue identified was the design or the interface of the devices, which raised many concerns among older adults. It was described by participants as uncomfortable or cumbersome [47,61,69,70,80,86,105,114,116,118]. For example, in 1 study [69], with regard to UNMTs (ie, motion and door sensors, smoke detectors, and fall detectors), the following information was provided:

> *Several older adults remarked on the light flashes coming from sensors.…She did not see its benefits and was also bothered by the light emitted from the sensors: "The watch kept falling out of her hand. It bothered her when she slept. She asked why the sensor light was flashing, why it was red, why this, and why that. It really drove me crazy. Then she said she did not want to wear it anymore, that it was useless" (Carer 5).*

Similarly, in another study [86], wearable activity trackers, in particular Garmin Vivofit2, elicited the following critique:

> *Nonusers described the band as plastic (ie, cheap and of bad quality), clunky, annoying, rigid, and uncomfortable: "It's very rigid. The design is poor. It collects water underneath. I end up having a really loose bracelet. Which could have some effect on accuracy. I don't know. I found it totally uncomfortable. It's really ugly."*

With regard to SARs, in 1 study [92], Hobbit was deemed "too big," and it was mentioned that "he gets very hot" and generates too much noise. In another study [116], the concern was about the excessive bulkiness of SARs, which could be dangerous if the older person lives in a small house with "potential presence of architectonic barriers." Moreover, the following statement was made: "Older adults' worries were also expressed in relation to the furniture or ornaments located inside their houses that could be damaged by robot movements." Other critical remarks were made about ETs, which were not user-friendly [84,91,106,121] and hard to understand [70,71,79,84,88,89,95]. In 1 study [84], older adults showed reluctance to use the wearable device Fitbit, which can monitor physical activity and vital signs:

> *…doubted their ability to manage the "complex" system.…For example, one senior said that "It sure takes a long time to use this Fitbit, I'm not sure it's worth it."…After using the devices,…seniors were more concerned with the output complexity of the wearable device, ie, the complexity of the measurements provided, expressed in incomprehensible data for them.*

Another type of UNMT, an integrated smart home system [70], was criticized in similar terms:

> *One participant…perceived the usability of an ISHS very complicated and was very hesitant to continue*



XSL•FO
RenderX

> *using an ISHS, "I find it difficult to use this technology, so I am not very confident I could properly use them in my life."*

Moreover, the interaction with technology evoked in the older adults feelings of frustration [29,70,82,84,89,90,95,99,111], disappointment [86], doubts about their efficacy and anxiety [82,86,91], fear [80], and apprehension [79]. These feelings sometimes emerged because of the device's inaccuracy and/or errors [84,86,99].

Participants also complained about their lack of knowledge and experience with the technologies and mentioned that this is a barrier preventing them from using the technologies [29,82,88,89,118]. For example, in 1 study [88], interviewees believed that "not knowing how to operate" mHealth technologies, such as mobile phones, tablet computers, patient monitoring devices, and mobile apps, "kept older adults from using such technology." Some interviewees commented:

> *I think they're good, but I just don't know how to use, to work it. I just don't have the knowledge on how to work it.*

> *I think the first thing is the lack of technical know-how is one of the barriers.*

Another barrier posed by some kinds of ETs was their inaccessibility to users with disabilities, such as reduced visual acuity [105].

Lastly, participants reported a sense of annoyance because the technology use was dependent on power supply and the need to charge the battery (and not simply replace it with a new one) [70,91,106].

Shifting the focus to publications that reported the benefits of ETs in terms of usability, many participants judged the different types of ETs as easy to use, user-friendly [46,54,61,75,82,87,88,92,98,109,112,114], and intuitive to operate and understand [57,87,107]. In other cases, the technology was appreciated because it was comfortable [70,90], versatile [118], and accessible, as it was designed with and for older people [114]. One paper showed how participants appreciated the communication service of SARs, enjoying the cooperation with the robots (ORO, DORO, and CORO) [93]. In 1 study [68], the reassuring features that indicated the technology was working were found to be important by older adults.

Analysis of several publications revealed that older adults have identified strategies to address 2 key usability issues: inadequate design/interface and lack of knowledge/experience. First, to improve the design of some technologies, participants were pragmatic in their recommendations [54,81,87]. For example, they asked for simple instructions, fewer buttons, larger fonts, manageable sizes, and speech-activated tools [81]. Second, to compensate for the lack of knowledge about technology, some participants suggested adequate and practical training on how to use it [79,88], while others highlighted the importance of support, encouragement, and education, considering that older adults need slow and individual guidance while using technology [111]. Other participants suggested receiving tailored instructions through readable and understandable manuals or handouts that accompany technology devices [79,81,118].

### Means to Address Vulnerabilities

Our analysis showed that older adults' suggestions for addressing vulnerability were closely tied to the specific technologies examined. Four publications [66,79,88,113] described strategies to mitigate particular vulnerability dimensions, including PHV [79]; RV [113]; MV understood as privacy/confidentiality [66] and decisional autonomy/self-determination [66,79,113]; SCV, particularly regarding isolation/exclusion from social life [113]; and EV [88].

In 1 study [79], which involved UNMTs (wireless sensor network-based systems that collect a range of environmental and structural sensory information such as weight and blood sugar), participants expressed 2 contrasting views. Some participants prioritized autonomy and independence, arguing that "users should be allowed to turn the system off when desired," thereby emphasizing MV mitigation related to decision-making. Others prioritized safety, proposing time limits on deactivation or preventing users from turning the system off entirely, given its purpose of enabling prompt medical assistance. PHV containment took precedence, even at the cost of reducing individual autonomy.

In another study [66], which was also focused on UNMTs, older adults suggested mitigating privacy-related MV by "processing the information before it is transferred." To further address MV interpreted as decisional autonomy, older adults emphasized that they "should decide what information is collected and transferred and to whom" and that technologies should be tailored to self-determined needs.

Regarding RV, participants in 1 study [113], which reflected on the use of a CMT (tablet-based communication app), stressed the need to "destigmatize loneliness, making it 'OK to talk about it' (Elsie, 86)" and avoid exacerbating reduced agency or compromised personhood. For this purpose, they recommended to "(1) understand interests and backgrounds to identify interventions, (2) provide a list of options for people to choose from and experiment with, and (3) ensure activities that entail active involvement and afford opportunities for meaningful interaction within and across generations." These suggestions support decisional autonomy and call for the improvement of SCV, understood in a broad sense as the possibility of connection and interaction with others and specifically as the reduction of isolation and exclusion from social life, by fostering intergenerational relations.

Finally, in 1 study [88], to address EV, participants recommended providing mHealth tools free of charge:

> *I would consider this kind of technology only if it were given to me.*

> *Well, I tell you like this, if they are going to pay for me to use one, then I will use one for my health…*



XSL•FO
RenderX

## Discussion

### Main Findings

This review aimed to systematically examine the perceptions and experiences of healthy older adults exposed to CMTs, UNMTs, VRTs, and SARs, and to investigate the ethically related impact of such technologies on the vulnerabilities of these older adults.

MV emerged as the most frequent concern. Many older adults emphasized the potential negative impact of ETs, especially monitoring technologies (wearables and environmental sensors), on privacy in terms of confidentiality, data handling, and disclosure to third parties. Others noted positive implications for autonomy and self-determination, as these devices provide greater control over oneself and one's living environment.

RV was identified as the second most frequently mentioned vulnerability. ETs positively influenced relationships among older adults, care staff, caregivers, and family members by supporting independence and community living while also facilitating practical and emotional support. Concerns remained that technologies might replace human interaction, reduce face-to-face care, increase misunderstanding and mistrust, or exacerbate loneliness, and this is consistent with prior literature, particularly regarding SARs [23,122-125].

Comparing these findings with those of another systematic review on older adults with cognitive impairments [126], we observed similar ambivalence. ETs can mitigate existing vulnerabilities but can also create new and/or worse (already existing) vulnerabilities. This depends on both the specific type of technology used and the dimension of vulnerability considered.

Both cognitively impaired older adults and healthy older adults recognized the positive impact of ETs on PHV, particularly through the ability of UNMTs to monitor health changes, detect falls and emergencies, and identify intruders, confirming previous findings [41].

A similar ambivalence emerged with RV. Yet, unlike cognitively impaired individuals (who focused on PV, RV, and PHV), cognitively healthy older adults emphasized the negative impact of ETs on MV. UNMTs, in particular, appeared to interfere with decisional autonomy, jeopardize privacy (confidentiality), compromise privacy through control and surveillance, undermine integrity and identity, and cause stigmatization and infantilization. This aligns with concerns described in the so-called *Big Brother Syndrome* [23], that is, the perception of being constantly observed by monitoring systems and feeling bewildered by the presence of numerous devices, which invade personal space [125]. In general, the literature shows that older adults are often willing to accept some privacy infringements in favor of safety and other benefits provided [127], but they nonetheless expect ETs to be reliable and trustworthy [128] and often react negatively to the invasiveness, intrusiveness, and obtrusiveness of ETs [122].

In conclusion, older adults with cognitive impairments tend to accept more tradeoffs in exchange for even modest physical or psychological benefits, due to greater psychophysical fragility, whereas cognitively healthy older adults tend to prioritize general, age-independent values, most notably privacy and autonomy.

### Strengths and Limitations

The main strength of this work is that our study design is unique in that there are no other systematic reviews examining the relationship among ETs (taken as a whole), cognitively healthy older adults, and ethically related content, with a focus on vulnerability. Second, this review allows both a broader overview of the impact of ETs on older adults' vulnerabilities and a detailed analysis of how *specific types* of ETs affect *distinct* vulnerability dimensions. In other words, our research is highly informative with regard to both its *breadth* (ie, the number of ETs included) and *depth* (ie, the fine-grained scrutiny it enables on the specific challenges that every single family of ETs raises on distinct dimensions of older adults' vulnerabilities). Third, as a complement to another systematic review of qualitative studies [126], this review contributes to advancing a broader research agenda on aging that includes both cognitively healthy and cognitively impaired populations. Finally, by reviewing qualitative studies conducted with older adults, this research foregrounds their perspectives, including needs, desires, and concerns.

There are some limitations. First, the search was conducted 2 years ago and included only English-language articles, potentially limiting its temporal and linguistic scope. Second, some included studies may have inadvertently involved participants with cognitive disorders (eg, Alzheimer disease), as screening for such conditions was not always explicitly reported in the primary research. Third, as this is a systematic review rather than a trial, the findings are not generalizable. Nonetheless, by synthesizing all available translated studies on the topic, this work offers a broad overview and highlights the conceptual trends of potential relevance.

### Conclusions and Future Directions

This study offers a solid theoretical foundation for future research, especially on strategies for addressing vulnerability in the context of aged care. Understanding older adults' vulnerabilities, especially in relation to ETs, is essential for identifying areas that demand greater sensitivity and tailored care approaches. Privacy and data confidentiality concerns (risks of breaches and data misuse or abuse) are crucial when deploying monitoring technologies, especially those involving audio or video recording in intimate settings. Cost also remains a significant barrier. Many older adults express interest in adopting ETs, but financial constraints may hinder their ability to do so. Equally important are the psychological impacts that such technologies may provoke, including anxiety, frustration, irritation, and feelings of inadequacy. Both ethics-by-design solutions and psychosocial interventions may help mitigate these effects.

Future research would benefit from complementing this theoretical framework with empirical studies aimed at identifying domains where psychological, sociopolitical, and economic interventions are most needed. Qualitative



XSL•FO
RenderX

methodologies would be particularly useful to capture the perspectives of key stakeholders, including formal and informal caregivers, health care professionals, and policymakers.

## Acknowledgments

The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content of this manuscript.

## Funding

This research is part of the project “Emerging Technologies and Vulnerabilities in Aged Care” (ElderTech), funded by Fondazione Cariplo (Social Research: Science, Technology and Society; grant number: 2020 - 1322).

## Data Availability

Additional data are provided in Multimedia Appendix 4. The conceptual schemes generated during this study are available from the corresponding author upon reasonable request.

## Authors' Contributions

CG and VS originated the idea of conducting a systematic review of qualitative evidence exploring the relationship between emerging technologies and vulnerabilities in older adults without cognitive impairments. CG assisted in devising the search algorithms, cross-checked publication selection, analyzed the material, and revised the final manuscript. AF, VS, and MDL conducted the literature search, worked out most of the search and analysis methods used, and analyzed and synthesized the material. AF wrote most of the paper. MDL, GB, and VS contributed to the writing of parts of the manuscript. All authors read and approved the final manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Groups of organizing concepts for searching the literature and their associated database search terms.
[DOCX File, 22 KB - i-jmr_v15i1e69676_app1.docx ]

Multimedia Appendix 2
Search strings used for searching databases stratified by organizing concepts.
[DOCX File, 21 KB - i-jmr_v15i1e69676_app2.docx ]

Multimedia Appendix 3
Inclusion and exclusion criteria.
[DOCX File, 23 KB - i-jmr_v15i1e69676_app3.docx ]

Multimedia Appendix 4
List of included publications.
[DOCX File, 80 KB - i-jmr_v15i1e69676_app4.docx ]

Multimedia Appendix 5
Detailed characteristics of the included publications.
[DOCX File, 41 KB - i-jmr_v15i1e69676_app5.docx ]

Multimedia Appendix 6
Example of the Qualitative Analysis Guide of Leuven conceptual scheme.
[DOCX File, 22 KB - i-jmr_v15i1e69676_app6.docx ]

Multimedia Appendix 7
Ethically related impact of emerging technologies on older adults’ vulnerabilities as gleaned from the perceptions and experiences of the older adults.
[DOCX File, 53 KB - i-jmr_v15i1e69676_app7.docx ]



XSL•FO
RenderX

Multimedia Appendix 8
Quality assessment of the included articles (Critical Appraisal Skills Programme).
[DOCX File, 55 KB - i-jmr_v15i1e69676_app8.docx ]

Checklist 1
PRISMA checklist.
[PDF File, 85 KB - i-jmr_v15i1e69676_app9.pdf ]

## References


1. UN decade of healthy ageing: plan of action. World Health Organization. 2020. URL: https://www.who.int/publications/m/item/decade-of-healthy-ageing-plan-of-action [accessed 2025-12-23]
2. WHO's work on the UN decade of healthy ageing (2021–2030). World Health Organization. URL: https://www.who.int/initiatives/decade-of-healthy-ageing [accessed 2024-02-08]
3. Healthy ageing and functional ability. World Health Organization. 2020. URL: https://www.who.int/news-room/questions-and-answers/item/healthy-ageing-and-functional-ability [accessed 2024-02-08]
4. World development indicators: population dynamics. World Bank Group. URL: https://wdi.worldbank.org/table/2.1 [accessed 2025-12-23]
5. Genge C, McNeil H, Debergue P, Freeman S. Technology to support aging in place: key messages for policymakers and funders. Front Psychol 2023;14:1287486. [doi: 10.3389/fpsyg.2023.1287486] [Medline: 38078234]
6. Etkin K. The AgeTech Revolution: A Book about the Intersection of Aging and Technology: New Degree Press; 2022.
7. Rotolo D, Hicks D, Martin BR. What is an emerging technology? Res Policy 2015 Dec;44(10):1827-1843. [doi: 10.1016/j.respol.2015.06.006]
8. Fasoli A, Beretta G, Pravettoni G, Sanchini V. Mapping emerging technologies in aged care: results from an in-depth online research. BMC Health Serv Res 2023 May 23;23(1):528. [doi: 10.1186/s12913-023-09513-5] [Medline: 37221528]
9. Pajalic Z, de Sousa DA, Strøm BS, et al. Welfare technology interventions among older people living at home-A systematic review of RCT studies. PLOS Digit Health 2023 Jan;2(1):e0000184. [doi: 10.1371/journal.pdig.0000184] [Medline: 36812629]
10. Vichitvanichphong S, Talaei-Khoei A, Kerr D, Ghapanchi AH. Assistive technologies for aged care: comparative literature survey on the effectiveness of theories for supportive and empowering technologies. Information Technology & People 2018 Mar;31(2):405-427. [doi: 10.1108/ITP-03-2017-0090]
11. Active ageing: a policy framework. World Health Organization. 2002. URL: https://iris.who.int/items/46476232-b305-47b4-ab28-e3931607c70c [accessed 2025-12-23]
12. Bernardo J, Apóstolo J, Loureiro R, et al. eHealth platforms to promote autonomous life and active aging: a scoping review. Int J Environ Res Public Health 2022 Nov 29;19(23):15940. [doi: 10.3390/ijerph192315940] [Medline: 36498018]
13. Patient engagement. World Health Organization. 2016. URL: https://iris.who.int/items/86fdfc1c-b581-4caf-97ab-6439ffac7cfb [accessed 2025-12-23]
14. Elwyn G, Frosch D, Thomson R, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012 Oct;27(10):1361-1367. [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]
15. Kondylakis H, Kazantzaki E, Koumakis L, et al. Development of interactive empowerment services in support of personalised medicine. Ecancermedicalscience 2014;8:400. [doi: 10.3332/ecancer.2014.400] [Medline: 24567757]
16. Antonovsky A. The salutogenic model as a theory to guide health promotion. Health Promot Int 1996;11(1):11-18. [doi: 10.1093/heapro/11.1.11]
17. Eriksson M. The sense of coherence in the salutogenic model of health. In: Mittelmark MB, Sagy S, Eriksson M, et al, editors. The Handbook of Salutogenesis: Springer; 2017:91-96. [doi: 10.1007/978-3-319-04600-6_11]
18. Ollevier A, Aguiar G, Palomino M, Simpelaere IS. How can technology support ageing in place in healthy older adults? A systematic review. Public Health Rev 2020 Nov 23;41(1):26. [doi: 10.1186/s40985-020-00143-4] [Medline: 33292707]
19. Kim D, Bian H, Chang CK, Dong L, Margrett J. In-home monitoring technology for aging in place: scoping review. Interact J Med Res 2022 Sep 1;11(2):e39005. [doi: 10.2196/39005] [Medline: 36048502]
20. Orlofsky S, Wozniak K. Older adults' experiences using Alexa. Geriatr Nurs (Lond) 2022 Nov;48:247-257. [doi: 10.1016/j.gerinurse.2022.09.017]
21. Lodi Rizzini C, Maino F, De Tommaso CV. Ageing in place, healthy ageing: local community involvement in the prevention approach to eldercare. SI 2023 Jan 25;12:1-17. [doi: 10.17645/si.7438]
22. Read E, Woolsey C, Donelle L, Weeks L, Chinho N. Passive remote monitoring and aging in place: a scoping review. Can J Aging 2023 Mar;42(1):20-32. [doi: 10.1017/S0714980822000198] [Medline: 35912590]
23. Sundgren S, Stolt M, Suhonen R. Ethical issues related to the use of gerontechnology in older people care: a scoping review. Nurs Ethics 2020 Feb;27(1):88-103. [doi: 10.1177/0969733019845132] [Medline: 31113266]




XSL•FO
RenderX

24. Czaja S, Beach S, Charness N, Schulz R. Older adults and the adoption of healthcare technology: opportunities and challenges. In: Sixsmith A, Gutman G, editors. Technologies for Active Aging. International Perspectives on Aging: Springer; 2013, Vol. 9:27-46. [doi: 10.1007/978-1-4419-8348-0_3]
25. Saplacan D, Khaksar W, Torresen J. On ethical challenges raised by care robots: a review of the existing regulatory-, theoretical-, and research gaps. Presented at: 2021 IEEE International Conference on Advanced Robotics and Its Social Impacts (ARSO); Jul 8-10, 2021. [doi: 10.1109/ARSO51874.2021.9542844]
26. Schicktanz S, Schweda M. Aging 4.0? Rethinking the ethical framing of technology-assisted eldercare. Hist Philos Life Sci 2021 Aug 3;43(3):93. [doi: 10.1007/s40656-021-00447-x] [Medline: 34342739]
27. Predel C, Timmermann C, Ursin F, Orzechowski M, Ropinski T, Steger F. Conflicting aims and values in the application of smart sensors in geriatric rehabilitation: ethical analysis. JMIR Mhealth Uhealth 2022 Jun 23;10(6):e32910. [doi: 10.2196/32910] [Medline: 35737429]
28. Khaksar W, Lindblom DS, Bygrave LA, Torresen J. Robotics in elderly healthcare: a qualitative analysis of 20 recent European research projects. J Hum-Robot Interact 2025 Jun 30;14(2):1-38. [doi: 10.1145/3711936]
29. Wang S, Bolling K, Mao W, et al. Technology to support aging in place: older adults' perspectives. Healthcare (Basel) 2019 Apr 10;7(2):60. [doi: 10.3390/healthcare7020060] [Medline: 30974780]
30. Vandemeulebroucke T, Dierckx de Casterlé B, Welbergen L, Massart M, Gastmans C. The ethics of socially assistive robots in aged care. a focus group study with older adults in Flanders, Belgium. J Gerontol B Psychol Sci Soc Sci 2020 Oct 16;75(9):1996-2007. [doi: 10.1093/geronb/gbz070] [Medline: 31131848]
31. Puri A, Kim B, Nguyen O, Stolee P, Tung J, Lee J. User acceptance of wrist-worn activity trackers among community-dwelling older adults: mixed method study. JMIR Mhealth Uhealth 2017 Nov 15;5(11):e173. [doi: 10.2196/mhealth.8211] [Medline: 29141837]
32. Muñoz Esquivel K, Gillespie J, Kelly D, et al. Factors influencing continued wearable device use in older adult populations: quantitative study. JMIR Aging 2023 Jan 19;6:e36807. [doi: 10.2196/36807] [Medline: 36656636]
33. Lu W, Silvera-Tawil D, Yoon HJ, et al. Impact of the smarter safer homes solution on quality of life and health outcomes in older people living in their own homes: randomized controlled trial. J Med Internet Res 2025 Jan 22;27:e59921. [doi: 10.2196/59921] [Medline: 39608020]
34. Wu M, Xue Y, Ma C. The association between the digital divide and health inequalities among older adults in China: nationally representative cross-sectional survey. J Med Internet Res 2025 Jan 15;27:e62645. [doi: 10.2196/62645] [Medline: 39813666]
35. Betriana F, Tanioka T, Osaka K, Kawai C, Yasuhara Y, Locsin RC. Interactions between healthcare robots and older people in Japan: a qualitative descriptive analysis study. Jpn J Nurs Sci 2021 Feb 28;18(3):e12409. [doi: 10.1111/jjns.12409] [Medline: 33644983]
36. Fan W, Zhao R, Liu X, Ge L. Intelligent robot interventions for people with dementia: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2025 Mar 10;27:e59892. [doi: 10.2196/59892] [Medline: 40063933]
37. Åkerlind C, Martin L, Gustafsson C. eHomecare and safety: the experiences of older patients and their relatives. Geriatr Nurs (Lond) 2018 Mar;39(2):178-185. [doi: 10.1016/j.gerinurse.2017.08.004]
38. Sanchini V, Sala R, Gastmans C. The concept of vulnerability in aged care: a systematic review of argument-based ethics literature. BMC Med Ethics 2022 Aug 16;23(1):84. [doi: 10.1186/s12910-022-00819-3] [Medline: 35974362]
39. Howes J, Denier Y, Gastmans C. Electronic tracking devices for people with dementia: content analysis of company websites. JMIR Aging 2022 Nov 11;5(4):e38865. [doi: 10.2196/38865] [Medline: 36367765]
40. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: a review of possible methods. J Health Serv Res Policy 2005 Jan;10(1):45-53. [doi: 10.1177/135581960501000110] [Medline: 15667704]
41. Vandemeulebroucke T, Dierckx de Casterlé B, Gastmans C. The use of care robots in aged care: a systematic review of argument-based ethics literature. Arch Gerontol Geriatr 2018 Jan;74:15-25. [doi: 10.1016/j.archger.2017.08.014] [Medline: 28926749]
42. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100. [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
43. Dierckx de Casterlé B, Gastmans C, Bryon E, Denier Y. QUAGOL: a guide for qualitative data analysis. Int J Nurs Stud 2012 Mar;49(3):360-371. [doi: 10.1016/j.ijnurstu.2011.09.012] [Medline: 21996649]
44. Gastmans C. Dignity-enhancing nursing care: a foundational ethical framework. Nurs Ethics 2013 Mar;20(2):142-149. [doi: 10.1177/0969733012473772] [Medline: 23466947]
45. Greenwood DA, Young HM, Quinn CC. Telehealth remote monitoring systematic review: structured self-monitoring of blood glucose and impact on A1C. J Diabetes Sci Technol 2014 Mar;8(2):378-389. [doi: 10.1177/1932296813519311] [Medline: 24876591]
46. Göransson C, Eriksson I, Ziegert K, et al. Testing an app for reporting health concerns-experiences from older people and home care nurses. Int J Older People Nurs 2018 Jun;13(2):1-10. [doi: 10.1111/opn.12181] [Medline: 29210218]




XSL•FO
RenderX

47. Thomas L, Little L, Briggs P, McInnes L, Jones E, Nicholson J. Location tracking: views from the older adult population. Age Ageing 2013 Nov;42(6):758-763. [doi: 10.1093/ageing/aft069] [Medline: 23761455]
48. Byrne CA, Collier R, O'Hare GMP. A review and classification of assisted living systems. Information 2018;9(7):182. [doi: 10.3390/info9070182]
49. Stavropoulos TG, Papastergiou A, Mpaltadoros L, Nikolopoulos S, Kompatsiaris I. IoT wearable sensors and devices in elderly care: a literature review. Sensors (Basel) 2020 May 16;20(10):1-22. [doi: 10.3390/s20102826] [Medline: 32429331]
50. Lányi CS. Virtual reality in healthcare. In: Ichalkaranje N, Ichalkaranje A, Jain L, editors. Intelligent Paradigms for Assistive and Preventive Healthcare. Studies in Computational Intelligence: Springer; 2006, Vol. 19:87-116. [doi: 10.1007/11418337_3]
51. Brown JA. An exploration of virtual reality use and application among older adult populations. Gerontol Geriatr Med 2019;5:2333721419885287. [doi: 10.1177/2333721419885287] [Medline: 31723574]
52. Feil-Seifer D, Mataric MJ. Socially assistive robotics. IEEE Robot Automat Mag 2011;18(1):24-31. [doi: 10.1109/MRA.2010.940150]
53. Sharkey N, Sharkey A. The crying shame of robot nannies: an ethical appraisal. Interaction Studies 2010;11(2):161-190. [doi: 10.1075/is.11.2.01sha]
54. Bian C, Ye B, Hoonakker A, Mihailidis A. Attitudes and perspectives of older adults on technologies for assessing frailty in home settings: a focus group study. BMC Geriatr 2021 May 8;21(1):298. [doi: 10.1186/s12877-021-02252-4] [Medline: 33964887]
55. Boström M, Kjellström S, Björklund A. Older persons have ambivalent feelings about the use of monitoring technologies. TAD 2013;25(2):117-125. [doi: 10.3233/TAD-130376]
56. Cabrita M, Tabak M, Vollenbroek-Hutten MM. Older adults' attitudes toward ambulatory technology to support monitoring and coaching of healthy behaviors: qualitative study. JMIR Aging 2019 Mar 12;2(1):e10476. [doi: 10.2196/10476] [Medline: 31518252]
57. Choi YK, Thompson HJ, Demiris G. Internet-of-things smart home technology to support aging-in-place: older adults' perceptions and attitudes. J Gerontol Nurs 2021 Apr;47(4):15-21. [doi: 10.3928/00989134-20210310-03] [Medline: 34038251]
58. Chung J, Thompson HJ, Joe J, Hall A, Demiris G. Examining Korean and Korean American older adults' perceived acceptability of home-based monitoring technologies in the context of culture. Inform Health Soc Care 2017 Jan;42(1):61-76. [doi: 10.3109/17538157.2016.1160244] [Medline: 27100664]
59. Courtney KL. Privacy and senior willingness to adopt smart home information technology in residential care facilities. Methods Inf Med 2008;47(1):76-81. [doi: 10.3414/me9104] [Medline: 18213432]
60. Courtney KL, Demiris G, Rantz M, Skubic M. Needing smart home technologies: the perspectives of older adults in continuing care retirement communities. Inform Prim Care 2008;16(3):195-201. [doi: 10.14236/jhi.v16i3.694] [Medline: 19094406]
61. Davenport RD, Elzabadani H, Johnson JL, Helal A(, Mann WC. Pilot live-in trial at the GatorTech Smarthouse. Top Geriatr Rehabil 2007 Jan;23(1):73-84. [doi: 10.1097/00013614-200701000-00010]
62. Demiris G, Rantz M, Aud M, et al. Older adults' attitudes towards and perceptions of "smart home" technologies: a pilot study. Med Inform Internet Med 2004 Jun;29(2):87-94. [doi: 10.1080/14639230410001684387] [Medline: 15370989]
63. Demiris G, Oliver DP, Dickey G, Skubic M, Rantz M. Findings from a participatory evaluation of a smart home application for older adults. Technol Health Care 2008;16(2):111-118. [Medline: 18487857]
64. Demiris G, Hensel BK, Skubic M, Rantz M. Senior residents' perceived need of and preferences for "smart home" sensor technologies. Int J Technol Assess Health Care 2008;24(1):120-124. [doi: 10.1017/S0266462307080154] [Medline: 18218177]
65. Dermody G, Fritz R, Glass C, Dunham M, Whitehead L. Factors influencing community-dwelling older adults' readiness to adopt smart home technology: a qualitative exploratory study. J Adv Nurs 2021 Dec;77(12):4847-4861. [doi: 10.1111/jan.14996] [Medline: 34477222]
66. Elers P, Hunter I, Whiddett D, Lockhart C, Guesgen H, Singh A. User requirements for technology to assist aging in place: qualitative study of older people and their informal support networks. JMIR Mhealth Uhealth 2018 Jun 6;6(6):e10741. [doi: 10.2196/10741] [Medline: 29875083]
67. Holender A, Sutton S, De Simoni A. Opinions on the use of technology to improve tablet taking in >65-year-old patients on cardiovascular medications. J Int Med Res 2018 Jul;46(7):2754-2768. [doi: 10.1177/0300060518770578] [Medline: 29730949]
68. Holthe T, Halvorsrud L, Lund A. A critical occupational perspective on user engagement of older adults in an assisted living facility in technology research over three years. J Occup Sci 2020 Jul 2;27(3):376-389. [doi: 10.1080/14427591.2020.1758200]
69. Hvalič-Touzery S, Smole-Orehek K, Dolničar V. Exploring reciprocity in perceptions on telecare within the informal carer-care receiver dyad. Teorija in Praksa 2021;58(3):840-859. [doi: 10.51936/tip.58.3.840-859]
70. Jo TH, Ma JH, Cha SH. Elderly perception on the internet of things-based integrated smart-home system. Sensors (Basel) 2021 Feb 11;21(4):1284. [doi: 10.3390/s21041284] [Medline: 33670237]




XSL•FO
RenderX

71. Knowles B, Hanson VL. Older adults' deployment of 'distrust'. ACM Trans Comput-Hum Interact 2018 Aug 31;25(4):1-25. [doi: 10.1145/3196490]
72. Leikas J, Kulju M. Ethical consideration of home monitoring technology: a qualitative focus group study. Gerontechnology 2018 Apr 16;17(1):38-47. [doi: 10.4017/gt.2018.17.1.004.00]
73. Lie MLS, Lindsay S, Brittain K. Technology and trust: older people's perspectives of a home monitoring system. Ageing Soc 2016 Aug;36(7):1501-1525. [doi: 10.1017/S0144686X15000501]
74. Peek STM, Luijkx KG, Rijnaard MD, et al. Older adults' reasons for using technology while aging in place. Gerontology 2016;62(2):226-237. [doi: 10.1159/000430949] [Medline: 26044243]
75. Pol M, van Nes F, van Hartingsveldt M, Buurman B, de Rooij S, Kröse B. Older people's perspectives regarding the use of sensor monitoring in their home. Gerontologist 2016 Jun;56(3):485-493. [doi: 10.1093/geront/gnu104] [Medline: 25384761]
76. Portet F, Vacher M, Golanski C, Roux C, Meillon B. Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. Pers Ubiquit Comput 2013 Jan;17(1):127-144. [doi: 10.1007/s00779-011-0470-5]
77. Sánchez VG, Anker-Hansen C, Taylor I, Eilertsen G. Older people's attitudes and perspectives of welfare technology in Norway. J Multidiscip Healthc 2019;12:841-853. [doi: 10.2147/JMDH.S219458] [Medline: 31802884]
78. Stack E, King R, Janko B, et al. Could in-home sensors surpass human observation of people with Parkinson's at high risk of falling? An ethnographic study. Biomed Res Int 2016;2016:3703745. [doi: 10.1155/2016/3703745] [Medline: 26981528]
79. Steele R, Lo A, Secombe C, Wong YK. Elderly persons' perception and acceptance of using wireless sensor networks to assist healthcare. Int J Med Inform 2009 Dec;78(12):788-801. [doi: 10.1016/j.ijmedinf.2009.08.001] [Medline: 19717335]
80. Walsh K, Callan A. Perceptions, preferences, and acceptance of information and communication technologies in older-adult community care settings in Ireland: a case-study and ranked-care program analysis. Ageing Int 2011 Mar;36(1):102-122. [doi: 10.1007/s12126-010-9075-y] [Medline: 21654869]
81. Wild K, Boise L, Lundell J, Foucek A. Unobtrusive in-home monitoring of cognitive and physical health: reactions and perceptions of older adults. J Appl Gerontol 2008;27(2):181-200. [doi: 10.1177/0733464807311435] [Medline: 19165352]
82. Ehn M, Eriksson LC, Åkerberg N, Johansson AC. Activity monitors as support for older persons' physical activity in daily life: qualitative study of the users' experiences. JMIR Mhealth Uhealth 2018 Feb 1;6(2):e34. [doi: 10.2196/mhealth.8345] [Medline: 29391342]
83. Essén A. The two facets of electronic care surveillance: an exploration of the views of older people who live with monitoring devices. Soc Sci Med 2008 Jul;67(1):128-136. [doi: 10.1016/j.socscimed.2008.03.005]
84. Farivar S, Abouzahra M, Ghasemaghaei M. Wearable device adoption among older adults: a mixed-methods study. Int J Inf Manage 2020 Dec;55:1-14. [doi: 10.1016/j.ijinfomgt.2020.102209] [Medline: 32834339]
85. Hein Willius A, Torres Hidalgo M, Arroyo Zuñiga P, et al. An acceptability study of a personal portable device storing critical health information to ensure treatment continuity of home-dwelling older adults in case of a disaster. Patient Prefer Adherence 2019;13:1941-1949. [doi: 10.2147/PPA.S218232] [Medline: 31806942]
86. Kononova A, Li L, Kamp K, et al. The use of wearable activity trackers among older adults: focus group study of tracker perceptions, motivators, and barriers in the maintenance stage of behavior change. JMIR Mhealth Uhealth 2019 Apr 5;7(4):e9832. [doi: 10.2196/mhealth.9832] [Medline: 30950807]
87. Thilo FJ, Bilger S, Halfens RJ, Schols JM, Hahn S. Involvement of the end user: exploration of older people's needs and preferences for a wearable fall detection device - a qualitative descriptive study. Patient Prefer Adherence 2017;11:11-22. [doi: 10.2147/PPA.S119177] [Medline: 28053509]
88. Cajita MI, Hodgson NA, Lam KW, Yoo S, Han HR. Facilitators of and barriers to mHealth adoption in older adults with heart failure. Comput Inform Nurs 2018 Aug;36(8):376-382. [doi: 10.1097/CIN.0000000000000442] [Medline: 29742549]
89. Johnson A, Shukla N, Halley M, et al. Barriers and facilitators to mobile health and active surveillance use among older adults with skin disease. Health Expect 2021 Oct;24(5):1582-1592. [doi: 10.1111/hex.13229] [Medline: 34190397]
90. Ladin K, Porteny T, Perugini JM, et al. Perceptions of telehealth vs in-person visits among older adults with advanced kidney disease, care partners, and clinicians. JAMA Netw Open 2021 Dec 1;4(12):e2137193. [doi: 10.1001/jamanetworkopen.2021.37193] [Medline: 34870680]
91. van Houwelingen CT, Ettema RG, Antonietti MG, Kort HS. Understanding older people's readiness for receiving telehealth: mixed-method study. J Med Internet Res 2018 Apr 6;20(4):e123. [doi: 10.2196/jmir.8407] [Medline: 29625950]
92. Bajones M, Fischinger D, Weiss A, et al. Results of field trials with a mobile service robot for older adults in 16 private households. J Hum-Robot Interact 2020 Jun 30;9(2):1-27. [doi: 10.1145/3368554]
93. Bevilacqua R, Felici E, Cavallo F, Amabili G, Maranesi E. Designing acceptable robots for assisting older adults: a pilot study on the willingness to interact. Int J Environ Res Public Health 2021 Oct 12;18(20):1-9. [doi: 10.3390/ijerph182010686] [Medline: 34682433]
94. Coghlan S, Waycott J, Lazar A, Neves BB. Dignity, autonomy, and style of company: dimensions older adults consider for robot companions. Proc ACM Hum Comput Interact 2021 Apr;5(CSCW1):1-24. [doi: 10.1145/3449178] [Medline: 34308262]
95. Deutsch I, Erel H, Paz M, Hoffman G, Zuckerman O. Home robotic devices for older adults: opportunities and concerns. Comput Human Behav 2019 Sep;98:122-133. [doi: 10.1016/j.chb.2019.04.002]




XSL•FO
RenderX

96. McGlynn SA, Kemple SC, Mitzner TL, King CH, Rogers WA. Understanding older adults' perceptions of usefulness for the Paro robot. Proc Hum Factors Ergon Soc Annu Meet 2014 Sep;58(1):1914-1918. [doi: 10.1177/1541931214581400] [Medline: 31320791]
97. McGlynn SA, Kemple S, Mitzner TL, King CHA, Rogers WA. Understanding the potential of PARO for healthy older adults. Int J Hum Comput Stud 2017 Apr;100:33-47. [doi: 10.1016/j.ijhcs.2016.12.004] [Medline: 28943748]
98. Niemelä M, van Aerschot L, Tammela A, Aaltonen I, Lammi H. Towards ethical guidelines of using telepresence robots in residential care. Int J of Soc Robotics 2021 Jun;13(3):431-439. [doi: 10.1007/s12369-019-00529-8]
99. Pripfl J, Körtner T, Batko-Klein D, Hebesberger D, Weninger M, Gisinger C. Social service robots to support independent living: experiences from a field trial. Z Gerontol Geriatr 2016 Jun;49(4):282-287. [doi: 10.1007/s00391-016-1067-4] [Medline: 27220733]
100. Ostrowski AK, Harrington CN, Breazeal C, Park HW. Personal narratives in technology design: the value of sharing older adults' stories in the design of social robots. Front Robot AI 2021;8:716581. [doi: 10.3389/frobt.2021.716581] [Medline: 34651018]
101. Harrefors C, Axelsson K, Sävenstedt S. Using assistive technology services at differing levels of care: healthy older couples' perceptions. J Adv Nurs 2010 Jul;66(7):1523-1532. [doi: 10.1111/j.1365-2648.2010.05335.x] [Medline: 20497268]
102. Broadbent E, Tamagawa R, Patience A, et al. Attitudes towards health-care robots in a retirement village. Australas J Ageing 2012 Jun;31(2):115-120. [doi: 10.1111/j.1741-6612.2011.00551.x] [Medline: 22676171]
103. Welsh S, Hassiotis A, O'mahoney G, Deahl M. Big brother is watching you--the ethical implications of electronic surveillance measures in the elderly with dementia and in adults with learning difficulties. Aging Ment Health 2003 Sep;7(5):372-375. [doi: 10.1080/1360786031000150658]
104. Baric V, Andreassen M, Öhman A, Hemmingsson H. Using an interactive digital calendar with mobile phone reminders by senior people - a focus group study. BMC Geriatr 2019 Apr 23;19(1):116. [doi: 10.1186/s12877-019-1128-9] [Medline: 31014276]
105. Fisher K, Easton K. The meaning and value of digital technology adoption for older adults with sight loss: a mixed methods study. Technology and Disability 2019;30(4):177-184. [doi: 10.3233/TAD-180205]
106. Wilson-Nash C, Tinson J. 'I am the master of my fate': digital technology paradoxes and the coping strategies of older consumers. Journal of Marketing Management 2022 Feb 12;38(3-4):248-278. [doi: 10.1080/0267257X.2021.1945662]
107. Ehn M, Johansson AC, Revenäs Å. Technology-based motivation support for seniors' physical activity: a qualitative study on seniors' and health care professionals' views. Int J Environ Res Public Health 2019 Jul 8;16(13):1-20. [doi: 10.3390/ijerph16132418] [Medline: 31288398]
108. Daniele K, Marcucci M, Cattaneo C, Borghese NA, Zannini L. How prefrail older people living alone perceive information and communications technology and what they would ask a robot for: qualitative study. J Med Internet Res 2019 Aug 6;21(8):e13228. [doi: 10.2196/13228] [Medline: 31389341]
109. Peek STM, Luijkx KG, Vrijhoef HJM, et al. Understanding changes and stability in the long-term use of technologies by seniors who are aging in place: a dynamical framework. BMC Geriatr 2019 Aug 28;19(1):236. [doi: 10.1186/s12877-019-1241-9] [Medline: 31462214]
110. Huang TY, Huang C. Elderly's acceptance of companion robots from the perspective of user factors. Univ Access Inf Soc 2020 Nov;19(4):935-948. [doi: 10.1007/s10209-019-00692-9]
111. Kärki A, Sallinen M, Kuusinen J. How to live independently with or without technology? Stud Health Technol Inform 2015;217:306-310. [Medline: 26294489]
112. Tsai HYS, Shillair R, Cotten SR, Winstead V, Yost E. Getting grandma online: are tablets the answer for increasing digital inclusion for older adults in the U.S.? Educ Gerontol 2015;41(10):695-709. [doi: 10.1080/03601277.2015.1048165] [Medline: 26877583]
113. Barbosa Neves B, Waycott J, Maddox A. When technologies are not enough: the challenges of digital interventions to address loneliness in later life. Sociol Res Online 2023 Mar;28(1):150-170. [doi: 10.1177/13607804211029298]
114. Neves BB, Mead G. Digital technology and older people: towards a sociological approach to technology adoption in later life. Sociology 2021 Oct;55(5):888-905. [doi: 10.1177/0038038520975587]
115. Park YH, Chang HK, Lee MH, Lee SH. Community-dwelling older adults' needs and acceptance regarding the use of robot technology to assist with daily living performance. BMC Geriatr 2019 Aug 5;19(1):208. [doi: 10.1186/s12877-019-1227-7] [Medline: 31382887]
116. Pigini L, Facal D, Blasi L, Andrich R. Service robots in elderly care at home: users' needs and perceptions as a basis for concept development. Technology and Disability 2012;24(4):303-311. [doi: 10.3233/TAD-120361]
117. Chang CP, Lee TT, Mills ME. Experience of home telehealth technology in older patients with diabetes. Comput Inform Nurs 2017 Oct;35(10):530-537. [doi: 10.1097/CIN.0000000000000341] [Medline: 28291156]
118. Vaportzis E, Clausen MG, Gow AJ. Older adults perceptions of technology and barriers to interacting with tablet computers: a focus group study. Front Psychol 2017 Oct 4;8:1687. [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
119. ISO 9241-11:2018: ergonomics of human-system interaction-part 11: usability: definitions and concepts. International Organization for Standardization. URL: https://www.iso.org/standard/63500.html [accessed 2026-01-15]




XSL•FO
RenderX

120. Chung J, Demiris G, Thompson HJ. Ethical considerations regarding the use of smart home technologies for older adults: an integrative review. Annu Rev Nurs Res 2016;34(1):155-181. [doi: 10.1891/0739-6686.34.155] [Medline: 26673381]
121. Harris MT, Rogers WA. Developing a healthcare technology acceptance model (H-TAM) for older adults with hypertension. Ageing Soc 2023 Apr;43(4):814-834. [doi: 10.1017/s0144686x21001069] [Medline: 37007645]
122. Ienca M, Wangmo T, Jotterand F, Kressig RW, Elger B. Ethical design of intelligent assistive technologies for dementia: a descriptive review. Sci Eng Ethics 2018 Aug;24(4):1035-1055. [doi: 10.1007/s11948-017-9976-1] [Medline: 28940133]
123. Lehoux P, Grimard D. When robots care: public deliberations on how technology and humans may support independent living for older adults. Soc Sci Med 2018 Aug;211:330-337. [doi: 10.1016/j.socscimed.2018.06.038]
124. Tan SY, Taeihagh A. Governing the adoption of robotics and autonomous systems in long-term care in Singapore. Policy and Society 2021 Apr 3;40(2):211-231. [doi: 10.1080/14494035.2020.1782627]
125. Zhu J, Shi K, Yang C, et al. Ethical issues of smart home-based elderly care: a scoping review. J Nurs Manag 2022 Nov;30(8):3686-3699. [doi: 10.1111/jonm.13521] [Medline: 34806243]
126. Sanchini V, Fasoli A, Beretta G, Gastmans C. Emerging technologies and vulnerabilities in older adults with cognitive impairments: a systematic review of qualitative evidence. BMC Geriatr (forthcoming) . [doi: 10.1186/s12877-025-06792-x]
127. Ji YA, Kim HS. Scoping review of the literature on smart healthcare for older adults. Yonsei Med J 2022 Jan;63(Suppl):S14-S21. [doi: 10.3349/ymj.2022.63.S14] [Medline: 35040602]
128. Pirzada P, Wilde A, Doherty GH, Harris-Birtill D. Ethics and acceptance of smart homes for older adults. Inform Health Soc Care 2022 Jan 2;47(1):10-37. [doi: 10.1080/17538157.2021.1923500] [Medline: 34240661]

## Abbreviations

**CMT:** conventional monitoring technique
**ESV:** existential or spiritual vulnerability
**ET:** emerging technology
**EV:** economic vulnerability
**MV:** moral vulnerability
**PHV:** physical vulnerability
**POV:** political vulnerability
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**PV:** psychological vulnerability
**QUAGOL:** Qualitative Analysis Guide of Leuven
**RV:** relational or interpersonal vulnerability
**SAR:** socially assistive robot
**SCV:** sociocultural vulnerability
**SPEV:** sociocultural, political, and economic vulnerability
**UNMT:** unconventional monitoring technique
**VRT:** virtual reality technique

*Edited by N Cahill, TDA Cardoso; submitted 05.Dec.2024; peer-reviewed by DS Lindblom, S Amadoru, YF Du; revised version received 30.Sep.2025; accepted 30.Sep.2025; published 19.Feb.2026.*

*Please cite as:*
*Fasoli A, De Luca M, Beretta G, Gastmans C, Sanchini V*
*Emerging Technologies and Vulnerabilities in Older Adults Without Cognitive Impairments: Systematic Review of Qualitative Evidence*
*Interact J Med Res 2026;15:e69676*
*URL: https://www.i-jmr.org/2026/1/e69676*
*doi:10.2196/69676*

© Annachiara Fasoli, Maria De Luca, Giorgia Beretta, Chris Gastmans, Virginia Sanchini. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 19.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Exploring Serious Games in Supporting Postnatal Depression: Narrative Review

Weijie Wang[1], MDes; Erika Penney[2], DClinPsy; Valerie Gay[3], PhD; Jaime Garcia[1], PhD

[1]School of Computer Science, Faculty of Engineering and IT (Information Technology), University of Technology Sydney, 15 Broadway, Ultimo, Sydney, Australia

[2]Graduate School of Health, Faculty of Health, University of Technology Sydney, Sydney, Australia

[3]School of Electrical and Data Engineering, Faculty of Engineering and IT (Information Technology), University of Technology Sydney, Sydney, Australia

**Corresponding Author:**
Jaime Garcia, PhD
School of Computer Science, Faculty of Engineering and IT (Information Technology), University of Technology Sydney, 15 Broadway, Ultimo, Sydney, Australia

## Abstract

**Background:** Postnatal depression (PND) is a clinical sign of sadness in certain individuals after childbirth. PND affects the mother, the baby, and the whole family. PND is now recognized as a public health concern worldwide. The global prevalence of PND is approximately 17.22%. However, less than half of those affected seek help, which means over 50% of PND cases are left untreated. Current reviews lack focus on digital interventions targeting parents in late pregnancy or postnatal stages. Existing studies prioritize symptom relief over fostering help-seeking behaviors.

**Objective:** This study aims to identify what serious games have been applied to support the treatment or help-seeking of PND and what gaps are still left.

**Methods:** Eligibility criteria for this review included full-text papers from 2015 to 2024 from conferences or peer-reviewed journals that were relevant to serious games to support help-seeking behaviors for individuals with depression. Seven research databases and publisher repositories were used. The final search was conducted in March 2025, and a thematic analysis was used to identify and organize recurring themes. As this review adopts a narrative approach, predefined eligibility criteria, a structured search strategy, and review by an interprofessional team were used to reduce selection bias.

**Results:** Only 2 studies related to PND were identified. After expanding the search string to depression, 13 studies were included in this review, and the studied games were divided into 3 help-seeking categories: promoting knowledge, reducing stigma, and raising awareness. This review identified that gamification, educational messages, and supportive character interactions could enhance engagement, build coping skills, and promote help-seeking in a practical, parent-friendly format. Nonetheless, this paper is limited by the reliance on depression literature due to scarce PND-specific studies, the quality of included studies, the exclusion of non-English language publications, and the use of common but select academic databases. These factors may affect generalizability but also serve to highlight critical gaps for future research and targeted intervention design.

**Conclusions:** There is a dearth of studies directly related to PND. Existing games commonly use narrative storytelling and interactive scenarios to promote empathy, correct misconceptions, and encourage help-seeking in broad depression. However, few are designed specifically for new parents, whose unique needs—such as time constraints—make mobile platforms the most suitable format for effective engagement. The authors propose that the future interprofessional codevelopment of a mobile serious game tailored to new parents would address the intervention and literature gaps identified in this review. It is argued that key design elements should include an emotionally engaging narrative, meaningful player choices, real-life parenting scenarios, calming visuals, and accessible, low-pressure gameplay. This review contributes to the progression of serious game research, with a focus on addressing the needs of an often underserved and undertreated PND population.

***(Interact J Med Res 2026;15:e70777)*** doi:10.2196/70777

**KEYWORDS**

serious game; postnatal depression; digital health; mobile phone; perinatal education; mental health



XSL•FO
RenderX

## Introduction

### Postnatal Depression: Definition, Consequences, and Treatment

Postnatal depression (PND) is a clinical diagnosis characterized by sadness after the birth of a child [1]. Most cases of PND start between 4 and 8 weeks after childbirth [2]. However, some research suggests that health care providers should screen for PND for up to 3 years after giving birth. This finding is in light of evidence that PND can develop after the traditional 6-month mark in some cases [3]. PND can result in various adverse maternal outcomes, impacting physical health, psychological well-being, interpersonal relationships, and risky behaviors [4]. Moreover, women who experience PND face an increased likelihood of experiencing clinical depression in the future, not just following repeated pregnancies but also throughout other periods of their lives [5].

PND affects not only the mother but the baby and family as a whole [6]. A mother experiencing PND may have difficulty bonding with her infant, providing proper care such as breastfeeding, and playing other maternal roles [1]. The potential consequences of this situation may extend to the child's physical, cognitive, language, emotional, social, and behavioral development and even impact the dynamic within the family unit over an extended period of time [7].

Perinatal mental health issues are now recognized as a significant public health concern all over the world. The global prevalence of PND was found to be approximately 17.22% in the largest meta-analysis of PND to date [8]. It is estimated that perinatal mental health disorders cost the Australian economy Aus $877 million per year (a currency exchange rate of Aus $1=US $0.66 was applicable) [9], the United States economy US $14.20 billion per year [10], and the United Kingdom £8.10 billion each year (a currency exchange rate of UK £1=US $1.34 was applicable) [11].

### Health-Seeking Behaviors and PND

However, less than half of those affected seek help, which means over 50% of PND is left untreated [12]. The barriers to help-seeking for PND are various, including social and cultural stigma, lack of knowledge and awareness, and lack of accessibility to the health care system [13]. As highlighted, PND can have a detrimental effect on the entire family. Thus, further research is needed on the most effective ways to empower expectant new parents to seek help earlier and maintain positive mental health, ensuring this condition does not remain a silent struggle.

### Digital Interventions and Management of PND

Digital games are interactive, computer-based experiences in which players interact with a virtual environment through input devices such as keyboards, controllers, or touchscreens. They range from simple, text-based applications to complex, immersive worlds powered by advanced graphics and artificial intelligence. Digital games can be played on multiple platforms, including personal computers, gaming consoles, smartphones, and virtual reality (VR) systems.

Game development is a complex process that often leverages specialized tools to streamline workflows and enhance functionality. Two critical components in this ecosystem are middleware solutions and game engines. Middleware provides developers with prebuilt modules to handle specific functionalities, while game engines offer comprehensive frameworks for building games.

Serious games play a crucial role in various domains by blending entertainment with purposeful outcomes, making them valuable tools for education, training, health care, and social awareness [14]. Unlike traditional video games designed solely for recreation, serious games leverage interactive and immersive elements to enhance learning, skill development, and behavioral change. Engaging users in simulated environments provides a risk-free space for experimentation and decision-making, leading to deeper comprehension and improved information retention. This kind of approach makes them particularly effective in fields such as education, where gamified learning experiences can make complex subjects more accessible and engaging for students. In professional training, serious games are widely adopted for skill development and simulation-based learning [15]. Industries such as health care, aviation, and military training use serious games to replicate real-world scenarios, allowing trainees to practice in a controlled setting. For example, medical simulation games help doctors and nurses refine surgical techniques and patient care by providing a safe space rather than performing procedures on real patients [16]. Similarly, corporate training programs integrate serious games to make the traditional learning experience more engaging and dynamic, thereby improving employee knowledge retention and job performance [17].

Beyond education and training, serious games significantly impact mental health and social awareness. Games designed to address depression can provide therapeutic benefits and be valuable and effective in reducing symptoms of depressive disorders [18,19]. Additionally, serious games are used to promote awareness about global challenges, such as climate change and poverty, by immersing players in experiences that foster empathy and drive real-world action [20]. By engaging storytelling, decision-making challenges, and interactive problem-solving, serious games go beyond traditional education and advocacy methods, making complex issues more relatable and actionable for diverse audiences.

In conclusion, serious games are an essential innovation in digital technology, bridging the gap between entertainment and meaningful engagement. Their ability to educate, train, and raise awareness while keeping users involved makes them a powerful tool across various industries. As technology continues to evolve, the potential for serious games to drive positive change in education, health care, and social development will only continue to grow. This narrative review aims to provide a comprehensive and interpretative narrative synthesis of the existing literature on serious games and digital interventions for PND. The goal is to critically evaluate current trends and gaps to explain the need for targeted research in this essential but often overlooked area.



XSL•FO
RenderX

### Knowledge Gap and Objective

There is literature from the past 10 years reviewing digital interventions for symptoms of PND. In 2018, van den Heuvel et al [21] published a systematic review of the current literature on eHealth developments in pregnancy to assess this new generation of perinatal care. Studies that reported using eHealth during prenatal, perinatal, and postnatal timeframes include 71 studies covering 6 domains. These domains are information and eHealth use (eg, patients' use of the internet for pregnancy information [22]), lifestyle factors such as gestational weight gain, exercise, and smoking cessation (eg, technology-supported diet and lifestyle interventions [23]), gestational diabetes (eg, telemedicine for diabetes in pregnancy [24]), mental health (eg, therapist-supported internet-based cognitive behavior therapy among postnatal women [25]), low- and middle-income countries (eg, mobile health [mHealth] interventions for prenatal, birth, and postnatal period in low- and middle-income countries [26]), and telemonitoring and teleconsulting (eg, wireless antepartum maternal-fetal monitoring [27]). There were 16 studies related to screening and treatment for PND (van den Heuvel et al [21], 2018).

Hussain-Shamsy et al [28], 2020, aimed to understand the extent, range, and nature of mHealth tools for prevention, screening, and treatment of perinatal depression and anxiety in order to identify gaps and inform opportunities for future work. Compared to the van den Heuvel et al [21] review, it targeted interventions delivered through mobile phones, including apps and text message–based interventions with prevention, screening, and treatment purposes. Tools were for prevention (10/22, 45%), screening (6/22, 27%), and treatment (6/22, 27%). Interventions delivered included psycho-education (16/22, 73%), peer support (4/22, 18%), and psychological therapy (4/22, 18%); however, interventions that started in pregnancy and continued into the postnatal period were rare (2/22, 9%).

In the same year, Dosani et al [29] investigated existing uses of mobile phone technologies for perinatal depression in low- and middle-income countries, finding improved depressive symptoms after the interventions. Similarly, Wan Mohd Yunus et al [30], 2022, found positive outcomes for digitalized cognitive behavioral therapy (CBT) interventions for depression symptoms during pregnancy. However, it should be noted that the studies within this systematic review demonstrated relatively high risks of bias and some missing outcome data. Lau et al [31] (2022) also conducted a review synthesizing 18 randomized controlled trials and demonstrating support for the effectiveness of digital CBT for perinatal psychology symptoms (ie, depression, anxiety, and stress symptoms) in high-income countries; however, again, the included studies were found to be of limited quality. It is argued that the literature is promising in its indicative efficacy for digital and mobile interventions for perinatal mental health; however, further robust, high-quality, larger sample size, and interprofessionally co-designed randomized controlled trial research between health care professionals and application developers is required in this field.

Beyond the more common digital and mobile interventions, VR has been explored for its utility in this field. In 2024, Fallon et al [32] conducted a scoping review on the use of VR to support parents during birth and in the first year postbirth in different settings, finding that across these studies, VR was found to be effective in improving both physiological and psychological outcomes. Furthermore, mothers reported positive experiences of using VR, which could indicate VR's acceptability in this population.

Literature has looked not only at interventions targeting symptoms but also at prevention. Some have assessed web- and mobile-based psychological interventions' role in preventing depression during the perinatal period [33], as well as reviews on the current state of diagnostic and screening apps for perinatal mental health [34]. These review studies highlight important considerations for future research, including the need for comprehensive digital assessment tools, ensuring data protection and safety of the intended app use, and improving data sharing features between users and health care professionals for timely support.

Other prevention-focused research has aimed at increasing partner support (Pilkington et al [35], 2015). Several prevention studies that include a partner component have demonstrated some benefits [35]. However, not all of these interventions were delivered to both mothers and fathers, and research evaluating their effects on paternal mental health is lacking. Thus, it is argued that future research needs to focus more on developing active interventions for both partners.

In the past decade, numerous reviews have focused on digital interventions for perinatal mental health, particularly PND. Each of these reviews had distinct aims, ranging from examining eHealth developments during pregnancy [21] to evaluating the effectiveness of mHealth tools [28] and digital CBT interventions [30]. Others, such as Fallon et al [32], explored emerging technologies such as VR for supporting parents during the postnatal period. These reviews covered a variety of tools and populations, including both high- and low-income countries, and targeted different stages of perinatal care. Efficacious tools are important. However, those with perinatal mental health symptoms have demonstrated delays in seeking those interventions [36]. To the best of the authors' knowledge, there are limited review papers focused on how digital interventions can improve help-seeking behaviors in perinatal depression, highlighting a significant literature gap with important clinical and economic implications for undertreatment.

There is a paucity of research that explores how the games are designed for PND. However, there is a game-specific exploration of depression symptoms more generally, which is of relevance given that PND and major depression share identical symptoms except for onset [37]. In 2014, there were 2 relevant review papers. One is the systematic review by Fleming et al [38] of the evidence for serious games in the treatment or prevention of depression. The current data suggest that it is possible to develop serious games for depression, that young people are willing to try them, and that available adherence and impact data, whilst limited, are promising. The other systematic review is Li et al [39] systematically examining the effectiveness of game-based digital interventions for depression, investigating psycho-education and training, VR Exposure Therapy, exercise, and entertainment.



XSL•FO
RenderX

Psycho-education and training and VR Exposure Therapy were identified as the most popular types of game applications for depression. Given the demonstrated potential gains for global access, it becomes even more crucial to meet the needs of those reluctant to seek help through current methods.

To the best knowledge of the authors, there is only one relevant review from 2018 (Dias et al [40]). Dias et al [40] aimed to identify through a systematic review how gamification and serious games support depression treatment and identified two significant literature gaps, including the limited study of (1) the effectiveness of gamification and serious games for depression and (2) hazard assessment of the side effects of using gamification and serious games in depression treatment. The technologies identified in this review included mobile, computer, wearable, and web applications that were applied in gamification, serious games, VR, and speech analysis. This area has not been revisited since this 2018 review, as far as the authors are aware, yet in recent years, there has been an increasing interest in serious games for depression and health problems more broadly [41].

A systematic review of casual video games (CVGs) by Pine et al [42] provided a systematic review of the effects of CVGs on treating anxiety, depression, stress, and low mood. This work includes 9 CVGs, of which 6 studies aimed at reducing anxiety, 2 studies examined effects for depression, and 4 studies investigated the impact of CVGs on treating stress or low mood. CVGs are becoming increasingly popular, and people report playing them for various reasons, such as relieving stress and relaxing, which may provide some initial benefit for those who need additional support or are waiting to access more comprehensive treatment.

In 2021, a total of 4 relevant works on games for depression and border mental health concerns were published in Australia, Ireland, Austria, and Taiwan. King et al [43] narratively reviewed indie—or independent—games that address mental health, trauma, and grief, highlighting developers' difficulties due to insufficient game design and research studies when creating, developing, and evaluating serious games. Indie games were initially defined as games made on low budgets by small teams that were published outside more mainstream channels and used by larger companies. There are many challenges when developing serious games, including limited funding, small teams, a need for broad expertise, and limited examples for developing both games and studies. Serious mental health games that target those with mental illness directly face additional challenges, such as mental health disorder symptoms that can affect the present ability to engage with content. This challenge means developing mental health serious games presents unique challenges to developers, who must decide how to best address their intended audience.

Differing from previous work that targeted independent games, Kowal et al [41] focused on the mental health benefits associated with playing commercial video games to address symptoms of depression and anxiety. In light of the current research, we conclude that commercial video games show great promise as inexpensive, readily accessible, internationally available, practical, and stigma-free resources for mitigating some mental health issues in the absence of, or in addition to, traditional therapeutic treatments. Aside from the game target general public, Martinez et al [44] investigated serious games published between 2015 and 2020 for depression and anxiety in children and adolescents, with a new approach focusing on their applications: awareness, prevention, detection, and therapy. The results of this systematic review show that more awareness and detection games and games with awareness, prevention, detection, and therapy applications are needed. In addition, games for depression and anxiety should equally target all age ranges. For future research, developing and evaluating serious games should be standardized, and the games should always offer support while playing. Meanwhile, Yen and Chiu [45] conducted a review to explore the effectiveness of VR exergames in improving older adults' cognition and ameliorating depressive outcomes. This study suggests that VR exergames can potentially positively influence cognition, memory, and depression in older adults. VR exergames could be an interesting strategy for active aging and a good mental health status.

In 2022, there were 5 key sources. First, Abd-Alrazaq et al [46] aimed to assess the effectiveness of serious games in alleviating depression by summarizing and pooling the results of previous studies. Their findings indicate that exergames are as effective as active interventions, which are usually delivered and supervised by health care providers. Meanwhile, Ruiz et al [18] systematically reviewed the evidence and found video game-based interventions were valuable and effective in reducing symptoms of depressive disorders. Additionally, Kim et al [47] conducted a systematic review, finding that serious games were beneficial in reducing depression in older adults. Regardless of the study setting, serious games appeared to reduce depression. Particularly, serious games, including physical activities, had a significant impact on reducing depression.

Concerning combined depression and anxiety-related symptoms, Townsend et al [19] conducted a review into the effectiveness of gaming interventions for treating either depression or anxiety in individuals aged 12 - 25 years. Preliminary evidence suggests that gaming interventions may be an effective treatment for youth depression but not anxiety. Further research is warranted to establish the utility, acceptability, and effectiveness of gaming interventions in treating mental health problems in young people.

The fifth relevant work of this year was published by Chitale et al [48], reviewing the viability of games and VR for the assessment of anxiety and depression, finding that possible digital correlates or biomarkers of depression and anxiety could help researchers with their design. It is important to emphasize that to ensure safety, efficacy, and privacy at a health care standard. It is argued that game developers and researchers must collaborate interprofessionally with qualified mental health specialists. More clinical data is necessary to further evidence the effective use of video games or VR in assessment methods for anxiety and depression.

In 2023, Gliosci et al [49] systematically reviewed the current array of scientific research on video games used as a therapeutic intervention tool for depression. The latest research on the use of video games as depression treatment methods shows that



XSL•FO
RenderX

there are significant gaps in the sector, particularly across generations. The age group that is most affected by the condition and now consumes the most video games—adults aged 18 - 40 years—has had limited focus in the game and mental health literature. Given that depressed people tend to be more isolated, there may be untapped possibilities, such as the unique way that games can bring people together, that could be used in different types of interventions.

Most recently, in 2025, Gómez-León [50] systematically reviewed the current serious games designed to train emotional regulation skills in children and adolescents, finding that serious games can be effective, acceptable, and feasible for learning emotional regulation strategies and reducing symptoms related to depression, anxiety, and lack of impulse control.

As highlighted, the current reviews have several limitations and clinically significant research gaps that need to be addressed. Existing reviews rarely address digital interventions specifically targeting parents in late pregnancy or postnatal stages. Most studies focus on adolescents, older people, or the general population with depression. Given the unique physiological and psychological needs of a perinatal mental health group and the common reluctance to seek help, it is crucial to develop targeted digital games and help-seeking interventions for new parents, especially during late pregnancy and the postnatal period.

Current reviews focus primarily on symptom relief rather than fostering help-seeking behaviors. For new parents, it is essential to explore how digital games can educate and provide psychological support to help them recognize symptoms of depression and empower them to seek professional help. Crucially, the physiological and psychological changes that occur during pregnancy and the postnatal period differ from those experienced by the general population with depression, often influenced by hormonal shifts and changes in life roles [1]. Existing reviews lack a deep exploration of how these specific mechanisms impact the effectiveness of gamified interventions, and therefore, this will be a particular aim of this review. Game design must consider these unique psychological and physiological changes to support the mental health of new parents effectively.

Although serious games have shown promise in supporting mental health, few have been tailored to the unique needs of new parents facing PND. Existing interventions often overlook practical constraints such as time limitations and emotional vulnerability. Given this need, the current review aims to explore how game-based tools can be thoughtfully designed to provide emotional support, foster help-seeking, and offer accessible education during this critical period.

There is growing recognition that serious games could offer an accessible, flexible, and emotionally supportive solution for broad mental health problems. This narrative review provides a comprehensive and interpretative narrative synthesis of the existing literature on serious games and digital interventions for PND and, where data is limited, extends its review to related literature on depression, given its almost identical diagnostic profile with PND. The aim is to critically assess prevailing trends and gaps and to inform the need for targeted research in key, often overlooked, areas. By addressing these aspects, we aim to make future direction recommendations for research and game developments that could cater to the unique needs of parents in late pregnancy and postnatal stages, particularly in raising awareness of help-seeking behaviors toward PND. This review aims to highlight how digital games can be more effective in supporting this specific underserved population.

## Methods

### Search Strategy and Study Selection

The literature search was conducted across 7 research databases and publisher repositories: PubMed Central, the ACM Digital Library, Google Scholar, the IEEE Xplore Digital Library, Science Direct, the Springer Library, and Wiley. Keywords included combinations of major terms such as “serious games,” “digital games,” “postnatal depression,” and “help-seeking.”

The resulting search string is as follows and shown in Table 1: ((serious game OR serious games OR gamification OR gamified OR game design OR game based OR game-based OR gaming) AND (postnatal depression OR postnatal depression OR perinatal depression OR maternal depression OR post-birth depression)). For the detailed search strategy, please refer to Multimedia Appendix 1.

**Table .** Search string.

| Major terms | Synonyms |
|---|---|
| Postnatal depression | postnatal depression OR perinatal depression OR maternal depression OR post-birth depression |
| Serious game | serious games OR gamification OR gamified OR game design OR game based OR game-based OR gaming |

The search covered literature published between 2015 and 2024 and was limited to studies published in English. We chose to only include articles published after 2015, considering relevance and time efficiency.

Given the only 2 game research specific to PND, we also chose to include depression as an analogous disorder for inclusion because, according to the *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition) [51], major depressive disorder and PND share core diagnostic symptoms. Thus, the broadened search string is as follows: ((serious game OR serious games OR gamification OR gamified OR game design OR game based OR game-based OR gaming) AND (depression OR sadness OR depressive))

Initial search results were screened based on titles and abstracts by author WW. Then, full-text reviews were conducted by author WW and decided across the team, EP and JG, to



XSL•FO
RenderX

determine their relevance to the research objectives. Reference lists of key articles were also reviewed to identify additional sources. The screening process is shown in the Results section.

Studies were included if they discussed the design, application, or evaluation of serious games related to promoting help-seeking behaviors toward depression or perinatal mental health. Articles focusing solely on maternal health without mention of game-based interventions were excluded.

### Sufficiency Statement

The authors acknowledge that their disciplinary backgrounds and research interests in digital mental health and game design may have influenced the interpretation and prioritization of specific themes during the analysis. Author WW is a PhD candidate who has experience in interaction design, data analysis, and software development. Author EP is a senior lecturer in the faculty of health. She has published reviews and empirical research in the field of mental health. She is also a clinical psychologist practitioner working with vulnerable and mentally unwell clients and families, including clients with PND. Author JG is a senior lecturer in games development, an experienced software engineer, and a serious game designer working with diverse clients. A particular emphasis was placed on design elements that support emotional engagement and help-seeking behaviors, which aligns with the authors' focus on developing interventions for new parents experiencing PND. Additionally, the narrative nature of the review allows for interpretive flexibility, which, while beneficial for exploring emerging ideas, may introduce subjective bias in theme identification and synthesis. The selection of literature in English only and a focus on peer-reviewed sources may have also limited the inclusion of culturally diverse perspectives and nontraditional forms of evidence.

### Analysis Approach

To guide the synthesis and interpretation of findings across the selected literature, this review adopted a thematic analysis approach. Commonly used in qualitative research, thematic analysis is a systematic method for identifying, analyzing, and interpreting patterns of meaning—referred to as themes—within textual data [52]. It is particularly appropriate for narrative reviews, as it allows for the integration of diverse sources and offers a deeper understanding of complex issues across varied contexts. The analysis followed the 6-step process [53]. Initially, all included texts were read thoroughly by author WW to ensure familiarity with the content. From there, initial codes were generated inductively by author WW to capture recurring ideas, concepts, and relevant patterns. These codes were then reviewed by the team (EP and JG) and grouped into broader themes that reflected meaningful trends across the dataset. Each theme was refined and decided across the team (WW, EP, and JG) to ensure clarity, distinctiveness, and relevance to the review's aims. The final step was author WW weaving the themes into a cohesive narrative that highlighted their significance in relation to the central research questions.

By using thematic analysis, this review was able to go beyond a simple summary, offering a structured yet flexible framework to interpret findings. A thematic narrative approach was considered appropriate due to the heterogeneity of the sources, which varied in methodology, population focus, and design strategy. This flexibility allowed for a deeper exploration of how serious games function as educational and emotional tools, especially in the context of PND.

## Results

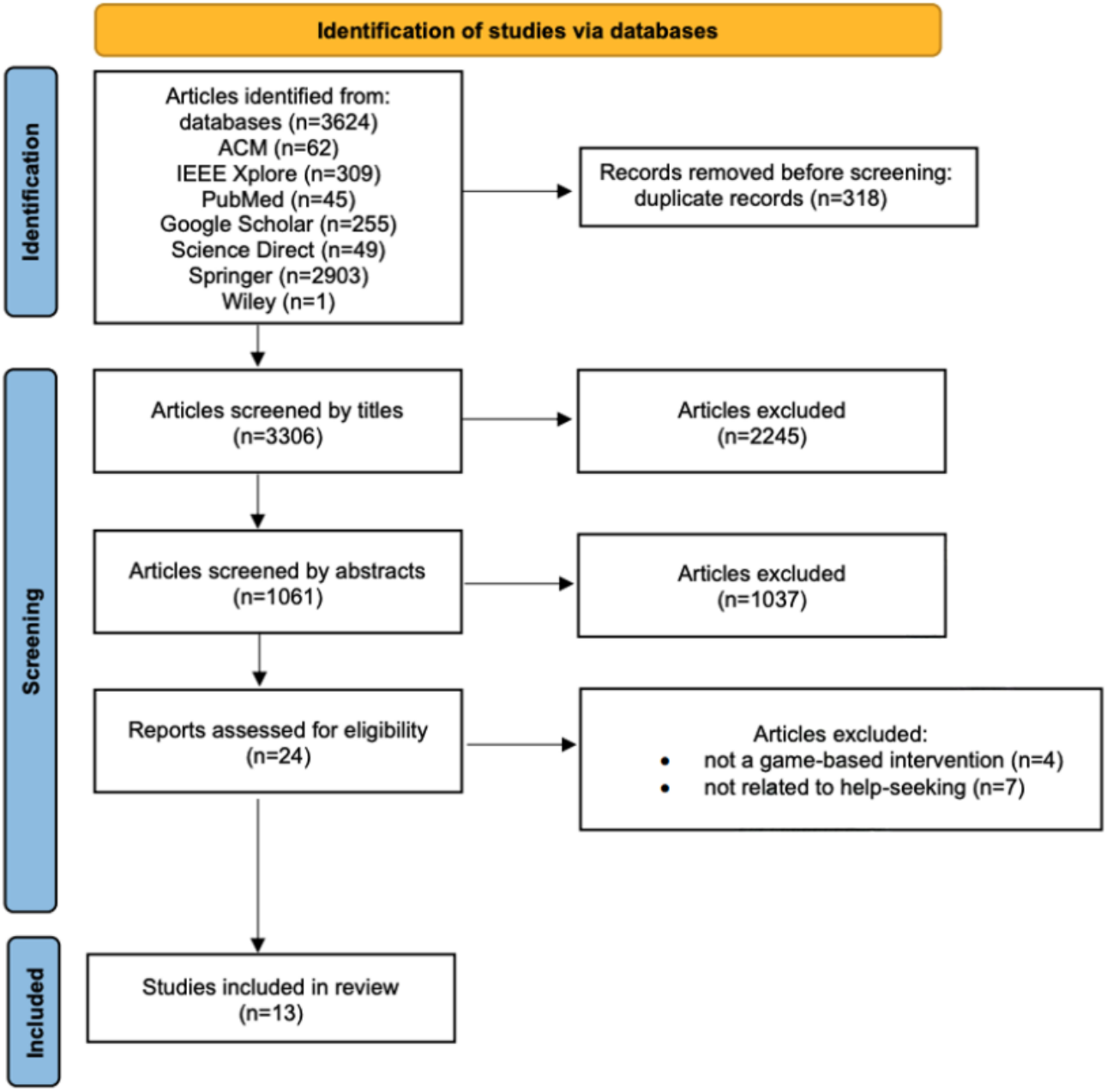
### Overview

There are 2 studies identified specifically from the search string PND, and 11 studies included from the search with depression. In total, 13 eligible studies were included in this review. There were 9 games published in academic articles, and the other 4 are from market research. In the past 5 years, there were 5/9 (55.6%) studies published. The screening process is shown in Figure 1. An overview of their characters is shown in Table 2.



XSL•FO
RenderX

**Figure 1.** PRISMA diagram. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.






XSL•FO
RenderX

**Table .** Summary of included studies.

| Name | Year | Country | Study design | Sample size | Intervention length |
|---|---|---|---|---|---|
| Above Water | 2016 | Canada | N/A[a] | N/A | N/A |
| Stigma-Stop | 2017 | Spain | Comparison study | 552 | N/A |
| Leadership Training in Mental Health Promotion (LMHP) | 2017 | Germany | Quasi-experimental study | 48 | 3 months |
| Moving Stories | 2019 | Netherlands | Cluster randomized controlled trial | 185 | 6 months |
| MindMax | 2020 | Australia | Participatory design | 40 | N/A |
| MANTRA | 2020 | United Kingdom | Cocreation process | 60 | N/A |
| Hellblade: Senua's Sacrifice | 2020 | United States | Two-condition, randomized design | 198 | 45 minutes |
| COEX-IST | 2022 | Brazil | N/A | N/A | N/A |
| PPDHero | 2024 | Bangladesh | Requirement elicitation study | 108 | N/A |

[a]N/A: the data were not clear or not mentioned in the original paper.

This section divides the 13 studied games included into 3 categories based on their perspectives on improving help-seeking behaviors: promoting knowledge, reducing stigma, and raising awareness.

## Promoting Knowledge

### *Overview*

An overview of the games that focus on promoting knowledge to help improve depression can be seen in Table 3.



XSL•FO
RenderX

**Table .** Games focused on promoting knowledge.

| Name | Year | Category | Aim | Audience | Game engine | Player mode |
|---|---|---|---|---|---|---|
| Above Water [54] | 2016 | Strategy games | This game informs people about the available strategies to cope with 2 types of anxiety disorders—generalized anxiety disorder and panic disorder. The game teaches players about existing treatments. | Game players | Physical cards and a mobile application | Multiple players |
| Moving Stories [55] | 2019 | RPGs[a] | This game aims to make adolescents have better mental health literacy and endorse fewer stigmatizing attitudes regarding depression. | Adolescents | Not specified | Single player |
| MindMax [56] | 2018 | Casual games | This game is to deliver psycho-educational modules and create a web-based community centering on well-being, AFL[b]. | Men aged 16 to 35 years who are interested in AFL or video games | Mobile phone | Single player |
| MANTRA [57] | 2020 | Casual games | This game is to increase maternal and child health resilience before, during, and after disasters. | Vulnerable low-literacy female audiences in rural Nepal | Mobile phone | Single player |
| PPDHero [58] | 2023 | Serious games | This game aims to screen PND[c] and provide education and support to new mothers to identify and manage symptoms and connect with health professionals. | New mothers | Mobile gamification app | Single player |

[a]RPG: role playing game.

[b]AFL: Australian Football League.

[c]PND: postnatal depression.

### *Above Water*

In 2016, Wehbe et al [54] released the educational game *Above Water* for anxiety. It is a digital or physical hybrid game to inform people about the available strategies to cope with 2 types of anxiety disorders, which are Generalized Anxiety Disorder and Panic Disorder. It is designed to inspire players to share their experiences and develop their narratives by teaching them existing treatments. *Above Water* is a game that players play with physical cards and digital devices, such as tablets and smartphones. The game is designed to be played with 4 players. They are required to achieve all 6 categories of life goals, including career, education, health, self-improvement, financial stability, and relationships, while managing anxiety. They need to select 4 life goals on their smartphone, which cannot be changed during the play session. On each player's turn, they will draw 1 card to face up, play, or discard. Anxiety cards can be discarded by using treatment cards. They can be taught deep breathing skills, yoga, and stretching poses using mobile applications. The digital implementation does not require downloading an app. It uses a simple HTML5-enabled website. This game has several strengths, particularly in raising awareness and encouraging social support through collaborative gameplay. Its focus on educating players about anxiety and introducing treatment options makes it helpful in fostering an open dialogue about mental health. However, the game falls short in targeting specific populations and providing tailored



XSL•FO
RenderX

interventions. Additionally, it lacks mechanisms for improving long-term behavior change and cultural adaptation, which are critical elements for effectively addressing the unique challenges faced by parents dealing with PND. It could be improved by incorporating specific content for new parents and addressing unique social and cultural barriers.

### *Moving Stories*

*Moving Stories* is a game-based school program for mental health literacy and stigma regarding depression published by Tuijnman et al [54] in 2019. It is developed by a professional game design company in collaboration with researchers and relevant stakeholders. Moving Stories includes three parts: (1) an introduction lesson, (2) a single-player, mobile, 3D video game, and (3) a contact session with someone with lived experience with a depressive disorder. The program focuses on three perspectives of mental health literacy: (1) recognizing the disorder, (2) having knowledge of help-seeking options and treatment available, and (3) first aid skills to support others who are developing or in a mental health disorder. The video game asks players to help a girl named Lisa, who has almost lost interest in everything. The players can do multiple things for Lisa to accumulate or decrease their relationship scores. As a bridge, the players are guided to share messages and discuss the reflection in a contact session with the school welfare coordinator or school counselor. This game demonstrates several strengths as a mental health intervention tool, particularly in improving mental health literacy, reducing stigma, and promoting help-seeking behaviors. By empowering players with knowledge, its approach to directly confronting myths through gameplay helps break down stigma. It encourages a more open conversation about mental health issues, which sets the foundation for increased understanding and empathy. It is crucial to tackle the stigma that new parents might face regarding PND. A key strength of the game lies in its focus on promoting help-seeking behavior. Through interactive scenarios that involve assisting the character Lisa and guiding her toward appropriate support, *Moving Stories* emphasizes the importance of recognizing mental health issues and seeking help. This game aligns well with the need to address the avoidance of help-seeking behaviors seen in PND, where new parents often struggle to ask for support due to stigma or lack of awareness. Incorporating role-playing elements demonstrates how and when to seek help, which is critical for empowering individuals to take action in real-life situations.

### *MindMax*

*MindMax* is an app that incorporates gamification, mini-games, and social connection to improve men's mental health and well-being, published by Cheng et al [56] in 2018. This game is based on the original well-being training program that the Australian Football League Players' Association offered players. Thanks to the collaboration with the University of Sydney and the University of Technology Queensland, it aims to deliver a portable, digital version of Australian Football League Players' Association's existing program. The design process includes 3 phases. The first phase is 6 participatory design workshops consisting of 3 stages: discovery, evaluation, and prototype to identify how best to frame the well-being concept discussed in *MindMax* and, more broadly, how to structure a mental health and well-being app for the intended audience. The second phase is knowledge translation. The third phase is user experience testing interviews. This work explained well why and how they implement the participatory design process. The core components of *MindMax* include two parts: (1) multiple educational modules, each around 10 minutes, and (2) a web-based community centering on well-being, sports, and video games. The good point in an educational context is that the audience does not care about what context is presented, but how it was presented. The tone of expression is very important. Regarding the social element, it suggested the app should enable rather than emulate, and sharing personal experiences with potential groups they are conscious of may need careful consideration.

One of the key strengths of *MindMax* is its incorporation of psycho-educational modules that provide users with mental health information in a clear and accessible manner. Each module is designed to be short (approximately 10 min), making it feasible for users with limited time, such as new parents with demanding schedules. These modules cover foundational topics such as mindfulness, emotional well-being, and basic coping skills. For new parents experiencing PND, such content can be particularly beneficial as it provides them with the tools they need to understand their emotional state and learn practical strategies for managing stress. Making these modules easily digestible helps ensure that even those unfamiliar with mental health concepts can engage with the material effectively. Besides, the mini-games and reward-based mechanics used in *MindMax* are particularly useful in maintaining user engagement, which might be a crucial factor for individuals who may be overwhelmed with responsibilities or have limited energy to devote to self-care. Using games as rewards helps create positive reinforcement for completing educational tasks, encouraging continued learning and interaction. For new parents, especially those struggling to make time for themselves, gamification can transform mental health support into something enjoyable and less intimidating, thus promoting consistent use of the intervention. The light-hearted games and achievement-based systems can provide much-needed moments of relaxation and enjoyment, breaking up the often stressful and exhausting routines of caring for a newborn.

### *MANTRA*

*MANTRA* is a serious mobile game that aims to reach a low-literacy audience in a low-resource setting with knowledge of maternal health, neonatal health, and geohazards. This work was published by Mueller et al [57] in 2020. The game mechanic is relatively simple in this game: picture matching with audio and visual feedback. The project's core goal is to tailor the knowledge to a local cultural context to fit specific settings with low literacy and resources. Throughout the design process, co-design and cocreation were guiding principles. The game design methodology builds on these principles and typical software development processes. The "drag-and-drop" is the only mechanism needed in the game. There is a touchscreen tutorial at first to help users who are not familiar with the screen interface. There are 3 modules in total, each containing 3 levels of complexity. The unsuccessfully answered question will be



XSL•FO
RenderX

pulled into an error pool and repeated up to 3 times throughout the level. The player will fail this level if all 3 times are answered incorrectly. *MANTRA* demonstrates several notable strengths in health education, particularly its accessibility for low-literacy users, cultural localization, and reinforcement of knowledge through repetition. It is designed for low-literacy users. It uses pictograms, audio prompts, and simple, intuitive interactions to ensure that even those with limited literacy skills can engage with and benefit from the game. This setting is particularly valuable for new parents from underserved communities who may lack access to formal health education, effectively addressing knowledge gaps. Second, the game's cocreation and cultural localization enhance its relevance and acceptance. Given that cultural context plays a significant role in the understanding and treatment of PND, incorporating cultural elements into the game helps to establish a deeper connection with users, reduce stigma, and increase engagement. Lastly, *MANTRA* uses repeated learning modules and immediate feedback, which helps reinforce knowledge retention. For new parents needing to master coping strategies and understand when to seek help, this method of repetition is crucial for effectively supporting their mental health outcomes. However, *MANTRA* also has some limitations. First, while using simple mechanics (such as drag-and-drop) makes the game accessible, it may not provide the depth of engagement needed for users facing significant mental health challenges. Second, the game lacks features that actively promote help-seeking behavior. For a digital game aimed at addressing PND, it is essential to include explicit pathways to professional support, such as information on local health care services or in-game guidance on seeking help, to bridge the gap between awareness and action. Lastly, the content of *MANTRA* primarily focuses on mothers without sufficient consideration of the role of fathers or nonbirthing parents in supporting maternal mental health.

### *PPDHero*

*PPDHero* is a novel web-based system published by Sara et al [58] in 2024 for new mothers to screen PND and provide them education and support to identify and manage symptoms and connect with health professionals if needed. The design process starts with an elicitation study, which involved 6 senior doctors, who were incorporated with surveys from the broader public. The system was developed based on interview and survey data analysis. It includes three modules: (1) relevant content management; (2) screening, assessment, and gamification; and (3) a treatment and support module to enable users to share resources, enroll in screening procedures and treatment plans, register for gamified activities, and involve trusted professionals in their postnatal journey. The gamification tasks, such as quizzes and exercises, are implemented in the screening and assessment session to enhance the assessment process. Its strengths lie in its comprehensive approach, integrating education, screening, and treatment in a single platform. Using gamification elements, such as quizzes and interactive activities, enhances user engagement and makes learning about PND more attractive. Additionally, the platform provides a personalized experience through tailored treatment plans based on screening results, ensuring that users receive targeted support to manage their mental health. However, while the platform provides education and support, it lacks features specifically aimed at reducing stigma, which is an important factor in encouraging new mothers to seek help.

## Reducing Stigma

### *Overview*

In Table 4, there is an overview of game-related components that contribute to depression and broad mental health illness by reducing stigma to the game players.

**Table .** Games focusing on reducing stigma.

| Name | Year | Category | Aim | Audience | Game engine | Player mode |
|---|---|---|---|---|---|---|
| Stigma-Stop [59] | 2017 | Simulation games | This game aims to reduce the stigma toward mental illnesses. | High school students | Unity3D | Single player |
| Leadership Training in Mental Health Promotion (LMHP) [60] | 2017 | Serious games | This game is to promote employee mental health and reduce mental illness stigma at work. | Managers | Digital game-based training program | Single player |
| Hellblade: Senua's Sacrifice [61] | 2020 | RPGs[a] | This game aiming to let players play video games featuring characters enduring mental illness may ultimately reduce stigma through transportation and identification. | Game players | Unreal Engine 4 | Single player |

[a]RPG: role playing game.



XSL•FO
RenderX

### *Stigma-Stop*

*Stigma-Stop* is a 3D serious game to reduce the stigma toward mental illness among high school students, published by Cangas et al [59] in 2017. It is specifically designed to reduce the stigma associated with mental health disorders by educating players about different mental health conditions, such as depression, schizophrenia, bipolar disorder, and panic disorder with agoraphobia. The game provides information about these disorders through interactive dialogues and scenarios where players must choose how to react to individuals with mental health issues. This kind of approach is effective in challenging and correcting misconceptions about mental illness, particularly among young people. The game presents players with scenarios of interacting with characters representing different mental health conditions. Players are asked how they perceive the characters' psychological state, whether they have experienced similar emotions, and how they can help the character. These choices, along with the feedback provided for correct and incorrect responses, allow players to learn more about mental health hands-on, engagingly. However, the game also has notable limitations. Its primary focus on adolescents and broader mental health disorders may make it less directly relevant for new parents dealing specifically with PND. Additionally, the game lacks practical coping strategies or therapeutic guidance, which limits its effectiveness as a support tool aimed at improving the help-seeking behaviors of PND.

### *Leadership Training in Mental Health Promotion (LMHP)*

*Leadership Training in Mental Health Promotion (LMHP)* is a digital game-based training program for managers to promote employee mental health and reduce mental illness stigma at work, published by Hanisch et al [60] in 2017. It applied various gamification components, such as providing a storyline and clear goals, including the capacity to overcome challenges by learning, giving feedback on performance, showing progress, and reinforcing learning by allocating points rather than simply providing badges for achievements and enabling competition between players, to facilitate an innovative and engaging learning experience. *LMHP* demonstrates several strengths that make it an effective tool for supporting mental health in the workplace and reducing stigma. The program's focus on mental health literacy, reducing stigma, and improving self-efficacy among managers makes it well-suited to foster a supportive work environment. The interactive simulation-based approach helps participants practice these skills in a safe environment, leading to higher engagement and better skill retention. While *LMHP* has strengths in terms of mental health literacy and stigma reduction, it focuses on the workplace environment and specifically targets managers. The scenarios and skills taught are tailored to workplace situations, which may not fully address new parents' specific emotional and psychological challenges in their home environment.

### *Hellblade: Senua's Sacrifice*

*Hellblade: Senua's Sacrifice* is an action-adventure game accurately portraying psychosis to reduce public mental health stigma, developed and published by Ninja Theory [61] in 2017. It aims to reduce the mental health stigma in 2 ways: by lowering stereotyping and limiting participants' desire for social distance. *Hellblade: Senua's Sacrifice* is set in an age of Vikings. Its story is about Senua's battles for the soul of her departed lover, Hela. Senua endures psychosis and must contend with her mental illness along with the challenges presented by her quest. The game blends gameplay mechanics and concepts such as puzzle solving, psychological horror, and melee combat. Voice acting is an integral part of the game, while its cut scenes combine motion capture by Melina Juergens and live-action performances by other actors.

The game's powerful and immersive portrayal of mental health struggles provides players with a unique opportunity to experience the emotional complexities associated with conditions such as psychosis. Through this kind of realistic depiction, new parents experiencing PND can see their struggles mirrored in the protagonist's experiences, helping them feel validated and understood. Second, the game's high-quality audio design plays a crucial role in enhancing the emotional depth of the experience. Using binaural audio to simulate auditory hallucinations, *Hellblade: Senua's Sacrifice* creates a profoundly immersive atmosphere that draws players into Senua's inner world. This kind of powerful audio experience can help new parents connect more profoundly with the emotional struggles portrayed, fostering empathy and offering a unique perspective on mental health challenges. Lastly, the game's development process involved collaboration with mental health experts and individuals with lived experience of psychosis. This process makes the portrayal of Senua's mental health struggles both vivid and respectful, adding credibility to the game's narrative.

## Raising Awareness

### *Overview*

In Table 5, there is an overview of how game-related components contribute to depression and broad mental health illnesses by raising awareness among game players.



XSL•FO
RenderX

**Table .** Games focused on raising awareness.

| Name | Year | Category | Aim | Audience | Game engine | Player mode |
|---|---|---|---|---|---|---|
| Depression Quest [62] | 2013 | RPGs[a] | This game aims to show other people with depression that they are not alone in their feelings and to illustrate to people who may not understand the illness the depths of what it can do to people. | Game players | Twine engine | Single player |
| Keep in Mind: Remastered [63] | 2018 | Simulation games | This game is meant to be a therapeutic experience for those who struggle with mental illness on a journey of reflection and emotional healing. | Those who struggle with mental illness or those who find themselves lost in the dark | Unity | Single player |
| Before I Forget [64] | 2020 | Simulation games | This game aims to raise awareness about mental health and communicate to the audience the struggles of people experiencing dementia. | Game players | Unity | Single player |
| Sea of Solitude [65] | 2021 | Simulation games | This game aims to reveal the themes of depression, loneliness, and hopelessness, how we unwittingly push people away, and how our actions can negatively affect others. | Game players | Unity | Single player |
| COEX-IST [66] | 2022 | Strategy games | This game is meant to increase mental health awareness and focus on self-care as a form of prevention. | People with a wide age range, from teenagers to adults | Not specified | Single player |

[a]RPG: role playing game.

### *Depression Quest*

*Depression Quest* is an interactive fiction game dealing with depression. It was developed by Zoe Quinn [62] using Twine (Interactive Fiction Technology Foundation) and published in 2013. This game aims to spread awareness by showing other people with depression that they are not alone in their feelings. *Depression Quest* is not an easy gaming experience because its goal is precisely to let you experience the difficulties of depressed people. It is a text-based game, and your choice will impact your ending. *Depression Quest* has a straightforward game structure. There are no animations or complex controls. As the story progresses, the player chooses different options to move the story in different directions. The part that *Depression Quest* comes closest to reality is that it tries to make people who do not have experience with depression understand that there are choices that do not exist for people who have depression. In the game, as the depressive condition deepens, the optimistic options are crossed out by the system with a red line—the player can only choose the closed, pessimistic options, and the depressive condition further deepens. There are fewer and fewer options to choose from in the game. It is a vicious circle and one that many people with depressive tendencies face.

This game has several strengths in its design elements. First, the game's narrative-driven depiction of depression provides a deeply empathetic look into the daily struggles of a person experiencing mental health challenges. Second, the game uses a unique mechanic where certain positive choices become restricted based on the protagonist's mental health status. This mechanic reflects the real-life constraints imposed by depression, such as the difficulty in making proactive decisions when feeling overwhelmed. For new parents experiencing PND,



XSL•FO
RenderX

this feature can help explain why even seemingly simple actions might feel impossible. It encourages self-compassion by showing that these limitations are part of the condition, not personal failings. Furthermore, as players engage with the narrative and take actions such as seeking help, they unlock new options, which effectively illustrate the empowerment that can come from actively addressing mental health challenges. Third, *Depression Quest* is also effective in raising awareness and fostering empathy among those who may not have firsthand experience with depression. By taking players through scenarios involving hopelessness, lack of motivation, and interpersonal struggles, this game mechanic could help partners and family members of new parents better understand the challenges their loved ones face. This improved understanding can lead to more supportive relationships and foster a compassionate environment. Finally, the game places a strong emphasis on encouraging help-seeking behavior. Players can pursue therapy or medication throughout the match, impacting the protagonist's situation and improving the choices. By showcasing how seeking help can positively affect mental health, *Depression Quest* actively promotes the idea that professional support is valuable and effective. New parents, who may feel hesitant to seek help due to stigma or uncertainty, could see the benefits of reaching out in this kind of game, which can help motivate them to pursue the support they need.

However, despite these significant strengths, the game has a key limitation. The intense focus on the struggles of depression, combined with the dark and sometimes overwhelming themes, may risk leaving players feeling disheartened without enough positive reinforcement or hope to balance the experience.

### *Keep in Mind: Remastered*

*Keep in Mind: Remastered* [63] is a heavy story-driven psychological indie game that follows Jonas, a man who struggles with grief, depression, and alcoholism, on a journey of reflection and emotional healing. One night, he awakens to a shadowy mirror world where beasts lurk, and stars do not shine. Lost and scared, Jonas must face the twisted beasts if he ever wishes to return home and learn the truth about his darkness. It was released by Little Moth Games and Akupara Games in March 2018. This game was created for those who struggle with mental illness or those who find themselves lost in the dark. This game is meant to be a therapeutic experience. *Keep in Mind: Remastered* has notable strengths that make it a compelling game for raising awareness about mental health challenges. For new parents, especially those experiencing PND, the portrayal of such internal struggles can help them feel seen and understood. This level of emotional storytelling can help reduce the stigma around PND by showing that struggling with mental health issues is not uncommon and that others share the negative thoughts and emotions they may be experiencing. Besides, the game uses symbolism effectively to represent the different facets of the protagonist's mental health challenges, such as dark creatures and eerie settings. This symbolism allows players to externalize their internal emotional struggles, making them more tangible and easier to understand. However, the game's focus on heavy and potentially triggering themes, lack of practical coping strategies, and absence of integrated help-seeking features might limit its effectiveness as a supportive intervention.

### *Before I Forget*

*Before I Forget* is a single-player exploration game developed and published by 3-Fold Games [64] in 2020. It is a short game that draws the player to a character experiencing dementia. She is housebound and unable to leave the flat. Only the windows could offer a view of the outside world, with its letterboxes and birdbaths. Otherwise, the exploration all happened in the mind as the character scrabbles to dredge artifacts from her past. In this game, players guide Sunita Appleby, who is a scientist with early-onset Alzheimer disease. As she interacts with now-unfamiliar objects in her house, some of her memories return. The game features puzzle-like elements, such as recalling where mementos have been left and asking players to guide Sunita to the bathroom when she perceives her house as a shifting maze. The game's use of empathetic storytelling and emotional depth provides a powerful connection to the struggles faced by those with mental health challenges. The game fosters understanding and reduces stigma around vulnerability and emotional struggle by allowing players to see the world through the protagonist's perspective. Additionally, the emphasis on relationships and emotional connections highlights the importance of support networks, encouraging players to lean on their loved ones during difficult times. However, the game also has notable limitations; the lack of practical coping strategies limits its effectiveness as a tool for managing day-to-day mental health issues, such as PND.

### *Sea of Solitude*

*Sea of Solitude* is an adventure video game developed by Jo-Mei Games and published by Electronic Arts [65]. The player controls a young woman named Kay, who endures such strong loneliness that her inner feelings of hopelessness, anger, and worthlessness turn to the outside, and she becomes a monster. As Kay, the player explores a seemingly empty flooded city and interacts with its scaly red-eyed creatures to reveal why she turned into a monster. Her emotions manifest into giant monsters standing in her way, trying to help but also destroy her. She needs to interact with and understand their underlying intentions to overcome the negative effects of those emotions. The game is an inner dialogue of a person trying to reconcile her shortcomings. It provides an insightful look at how mental illness devastates the lives of not just those it affects but also loved ones on the outside. Kay learns a lot about herself by understanding the value of listening, coming to terms with her flaws, and not just empathizing with her family but also accepting that a simple fix is not always possible. The game's powerful narrative effectively depicts the emotional struggles of loneliness, depression, and anxiety, allowing players to relate deeply to the protagonist's journey. Referring to this representation can help new parents normalize their emotions and reduce the stigma around experiencing mental health challenges during the postnatal period. The exploration mechanics and emotional expression throughout the game provide a meaningful way for players to reflect on their experiences. Additionally, the game's focus on empathy and understanding different characters' struggles helps to foster a



XSL•FO
RenderX

deeper connection between players and the themes of mental health. For new parents, this emphasis on empathy can reduce feelings of isolation and highlight the importance of reaching out for support, both for themselves and others. However, *Sea of Solitude* also has notable limitations in addressing PND effectively. The game lacks practical coping strategies that can be directly applied to real-life challenges. For new parents, having evidence-based tools and actionable advice would be instrumental in managing the day-to-day reality around PND. Without such strategies, the game may fail to deliver tangible help beyond emotional understanding. Moreover, the heavy themes of depression, anxiety, and loneliness may be overwhelming for some new parents, particularly those who are already vulnerable. Without sufficient pacing options or content warnings, the game risks triggering or intensifying negative emotions rather than alleviating them.

### *COEX-IST*

*COEX-IST* is a 3D interactive story decision-making game that aims to grow awareness about depression and social isolation after a pandemic. In 2022, Rodrigues et al [66] developed and published it. The game story is based on 25 undergraduate students' personal experiences. It is a third-person short game based on narratives and point-and-click to interact with the scene's objects and make decisions. The end of the narrative has 2 possible actions for game over: victory over depression or victory of depression. The story's resolution also involves tying up the loose ends of the climax and falling action. Each choice has impact consequences. Both possible ending scenes allow the player to experience and think about depression, growing awareness of this illness. The game's interactive decision-making process, emphasis on empathy, and reflective storytelling structure offer a deep and meaningful engagement that helps players connect with the content and understand the complexities of mental health challenges. First, the interactive decision-making feature allows players to experience the consequences of their actions realistically. This level of interactivity encourages new parents to see the potential impact of small, positive actions in their own lives, thereby fostering a sense of control over their mental well-being. This sense of agency is particularly important for new parents, who may often feel overwhelmed by the responsibilities and pressures of caring for a newborn. Second, the game's emphasis on mental health awareness and emotional empathy is effectively conveyed through storytelling. By representing depression as "Mr. Shadow," *COEX-IST* provides a powerful visualization of the weight and burden of mental illness. This representation helps players externalize and better understand their own struggles. It also reduces the stigma associated with these emotions by demonstrating that they are a shared experience that can be overcome with the right actions. This empathetic storytelling can provide comfort and connection for those who may feel isolated in their struggles. Lastly, the structured storytelling approach gives the narrative a clear and engaging emotional experience that mirrors the journey of dealing with mental health challenges. For new parents, seeing the protagonist's ups and downs depicted in a structured way can help normalize their own experiences. The 2 possible endings—either overcoming depression or being overtaken by it—underscore the importance of the player's choices and illustrate the ongoing nature of mental health management. This setting highlights that while setbacks may occur, progress is possible, which can provide hope and motivation to players during difficult times.

Many of the reviewed games (eg, *Depression Quest* and *Before I Forget*) aimed to raise awareness and build empathy through immersive experiences that replicate the emotional realities of those affected. These games often use narrative-driven storytelling to allow players to live through the struggles of a character facing depression. Through experiencing the emotional highs and lows, players are encouraged to empathize with the character's struggles. This emotional journey is argued to be crucial not only for players who are personally dealing with mental health challenges but also for their support networks—such as partners, family members, and colleagues—who may benefit from a better understanding of what their loved ones are going through. This approach can externalize internal emotional struggles, providing validation for players dealing with similar issues. The intention here is to not only entertain but offer a profound and empathetic experience that helps players feel less isolated.

Some reviewed games (eg, *MindMax* and *Stigma-Stop*) encouraged players to seek professional help for mental health concerns. By embedding choices in the game where players can seek therapy, use support resources, or talk to someone, these games guide players toward understanding the importance of professional help in managing mental health conditions.

Six games targeted adolescents and youth from 15 to 25 years, and one targeted younger children. Six games do not specifically target an age group, but board game players. Examples of these groups include 1 study for men who are interested in AFL games and 1 study for company managers.

Among these games, only 2 games targeted new mothers with a game intervention focused on promoting knowledge. Although few studies directly target new parents, there is arguably some overlap with the depressive symptoms faced by the other groups discussed above. However, new parents are often in a unique and vulnerable position, facing a combination of physical, emotional, and psychological challenges. They usually experience heightened vulnerability, exhaustion, and lack of time [67], which requires an intervention that is not only effective but also easily accessible and low-pressure. Many of the games targeting adults with depression do an excellent job of portraying the emotional complexities of mental health issues. Still, their intensity and focus on complex gameplay may not be suitable for new parents who are experiencing severe fatigue. It is argued that the new parent target audience would benefit more from short, accessible gameplay that can be easily integrated into their busy routines, and casual yet meaningful experiences that provide emotional validation and support.

Given the needs of the new parent target audience, mobile platforms are likely the most suitable choice. Mobile games can be played in short bursts, require minimal setup, and are highly accessible because most people carry smartphones. Mobile accessibility also allows parents to engage with the game whenever they find spare moments, which is crucial given their unpredictable and fatiguing schedules. However, a hybrid



XSL•FO
RenderX

solution that combines mobile and web-based access to ensure that new parents have multiple ways to engage with the game based on their convenience may also be valuable. A web-based version could provide a larger screen and more detail for those moments when new parents can access a computer, while the mobile version allows instant play. WebGL technology further reduces minimum system requirements and will enable games to be accessed from any electronic device. Hybrid accessibility also increases reach, as it accommodates different user preferences and ensures that players have multiple ways to interact with the content.

Console and VR platforms, while highly immersive, are arguably not ideal for most new parents. While VR today is widely "plug and play" and often requires little more than a connection ID to get started, the transition to becoming a parent, particularly in the early months, is marked by extreme time constraints, chronic sleep deprivation, and a constant state of divided attention [68]. New parents often navigate caring for an infant, are adjusting to unpredictable routines, and manage both emotional and physical exhaustion. Therefore, it is proposed that even simple steps such as clearing space, troubleshooting hardware, or remaining engaged in a virtual environment for more than a few minutes can become overwhelming for parents who might be at risk of cognitive and emotional overload.

## Discussion

### Overview

There is a dearth of evidence directly related to PND, which is only 2 studies in the past 10 years. After expanding the search to depression, we included 11 more. The included studies predominantly aligned with 3 overarching themes: the promotion of knowledge, the reduction of stigma, and the raising of awareness. A core intention of several of the reviewed games (eg, *Above Water* and *Moving Stories*) is to educate players about mental health and reduce the associated stigma. This is achieved by exposing players to the symptoms, challenges, and lived experiences of individuals who are dealing with various mental health conditions. The games use different approaches, such as presenting players with scenarios that require them to make supportive decisions, simulating real-life interactions with people facing mental health challenges, and providing informative feedback. The ultimate goal is to educate the players on mental health literacy, that is, what different symptoms mean, why they matter, and how to effectively respond. This type of education, delivered in an interactive, engaging way, generally helped to demystify mental health issues, correct misconceptions, and encourage empathetic attitudes. This is argued to play a crucial role in reducing stigma by promoting understanding and highlighting the commonality of these experiences, breaking down the barriers that often prevent open conversations about mental health.

### Game Mechanics

For depression management in general, several mechanisms are argued to be particularly impactful. Narrative storytelling, as seen in games such as *Hellblade: Senua's Sacrifice* and *Depression Quest*, may provide emotional engagement that helps players feel understood and validated by illustrating shared experiences of mental health struggles. Given the evidence that the recognition and acceptance of an individual's feelings and experiences promote mental well-being, seeing their experiences reflected in the storyline may create an emotional connection that could combat isolation. Choice-based interaction, where players make decisions that affect the storyline, may offer a sense of agency and sense of control, which is often decreased in those experiencing depression [69]. These decisions can illustrate how positive actions, such as seeking help or practicing self-care, might lead to better outcomes, potentially reinforcing a sense of empowerment for new parents. Simulated real-life scenarios could also play a role in helping players practice practical coping skills, such as managing stress or communicating with loved ones in a safe and controlled environment. Combined with resilience mechanics that highlight small wins, these mechanics may help maintain motivation during challenging times. Additionally, low-pressure gameplay is likely important to this target group, as it may reduce the cognitive and emotional needs of interaction, making the game more accessible to new parents who are experiencing fatigue and time constraints.

For educational purposes and stigma reduction, specific game mechanics may help build knowledge and empathy in accessible ways and minimally overwhelm the player. For example, educational micromessaging, as seen in "Match Emoji" and "Stigma-Stop," delivers brief and embedded content that increases awareness about mental health conditions. This mode of delivery may be especially valuable for new parents, who often benefit from concise, low-effort learning formats. Empathy-building mechanisms could also contribute to stigma reduction, an important function within PND given its high association with stigma [13]. Interactive elements that allow players to make supportive choices during character interactions may foster greater understanding and compassion. Role-playing as characters facing mental health challenges may help players consider these issues more personally, potentially encouraging self-compassion and empathy for others. Additionally, real-life scenario simulations may offer players opportunities to engage in realistic support-seeking situations, which could enhance a better understanding of seeking and receiving support.

Given the cultural, social, and personal barriers to seeking help for people with PND [13], certain game mechanics could play a valuable role in improving help-seeking behavior by reducing hesitancy and encouraging engagement with professional help. Choice-based interactions that include options for seeking therapy or speaking to a health care professional can illustrate the potential positive impact of these decisions and may help normalize the idea of reaching out. In-game scenarios that model conversations with health care professionals or trusted family members could also reduce anxiety by providing players with low-risk opportunities to observe or practice supportive dialogue. Additionally, incorporating positive reinforcement, such as visual or narrative cues that reflect improvements in well-being or relationships following help-seeking, may highlight the value of accessing help. For expectant parents who may experience PND after their baby is born, these mechanics may offer gentle encouragement and promote greater confidence



XSL•FO
RenderX

in taking the first step. As shown in Table 6, here is a comparison of game mechanics and different aims.

**Table .** Comparison of game mechanics and different aims.

| | Build knowledge | Build empathy | Build awareness | Reduce stigma |
|---|---|---|---|---|
| Narrative storytelling | ✓ | ✓ | | ✓ |
| Choice-based interaction | | ✓ | | ✓ |
| Educational micromessage | ✓ | | ✓ | |
| Character interaction | ✓ | ✓ | | |
| Real-life scenarios | ✓ | | ✓ | |
| Resilience mechanics | | | ✓ | |

## Principal Findings

To effectively support new parents coping with PND and to address the gaps identified in this review, the authors argue that future research should focus on the interprofessional co-development of a serious game with the following important design elements. First, we argue that a relatable and engaging story is important and should reflect both the struggles and victories that new parents face, helping them see their own experiences mirrored in the game. By doing this, the game can foster a deep sense of understanding, connection, and emotional validation, which may be crucial for reducing the isolation [13] often felt during PND. Allowing players to make meaningful choices that influence the direction of the game could give them a sense of agency and control to combat the low sense of agency [69] frequently found in the experience of depression. These decisions should reflect real-life challenges, and the consequences should demonstrate how positive actions, such as seeking help or practicing self-care, can improve well-being. This interactive element can empower new parents to feel more in control of their own mental health journey. Incorporating real-life scenarios is also a key element. These scenarios should include common parenting challenges, such as communicating effectively with a partner, managing stress, or seeking support. Practicing these skills in a safe, controlled game environment can help players build confidence, making it easier to apply what they have learned to real life.

Gamification elements such as badges, progress tracking, and daily challenges should be used to keep players motivated and committed, in line with a positive reinforcement approach to learning and improving intrinsic motivation [70]. These features add an element of fun and reward, encouraging players to engage consistently with the game. The game should also include educational messages that provide important information about PND. These messages could cover recognizing symptoms, understanding when to seek help, and learning practical self-care techniques. Delivering these messages in small, easily digestible portions ensures that players can learn without feeling overwhelmed, and it is in line with research findings that show benefits for adult learners who engage in mobile-based microlearning steps [71].

Character interactions are another essential design element. Players should have opportunities to interact with supportive characters, which can help build empathy and promote positive communication skills. These interactions can model supportive relationships and reinforce behaviors that new parents can use in their own lives to improve their support networks. Celebrating small wins is also important; highlighting moments such as taking time for self-care or successfully reaching out for help can reinforce resilience and make players feel that each positive step, no matter how small, is a significant achievement. Lastly, the game should be designed to be accessible and easy to engage with in short sessions. Low-pressure gameplay is important, as new parents often face exhaustion and have limited time [68]. The game should be flexible, allowing players to pick it up and put it down whenever they have a few moments, ensuring it serves as a helpful, nonstressful tool in their daily lives. This balance of emotional engagement, practical learning, and ease of access is proposed to create a supportive experience that can make a meaningful difference for new parents coping with PND.

## Recommendations

### *Overview*

Serving the aim of our proposed project, we recommend focusing on the following subset of game types and mechanics.

### *Narrative-Driven Storytelling*

Incorporate an engaging storyline that reflects the real-life experiences of new parents, including both struggles and successes. This mechanic helps players feel understood and connected, reducing feelings of isolation. The narrative should also include resilience themes to promote hope and perseverance.

### *Choice-Based Interactions*

Implement choices that directly influence the game's storyline, allowing players to make decisions that affect their mental health journey. This empowers players by giving them a sense of control, emphasizing the positive outcomes of seeking help or practicing self-care.

### *Real-Life Scenario Simulations*

Include scenarios that simulate everyday parenting challenges, such as communicating with a partner, managing stress, or seeking professional support. Practicing these skills in a game setting builds players' confidence and provides them with practical tools for real-life application.

### *Low-Pressure, Casual Gameplay*

Design the game for short, manageable sessions that can be played on a mobile device. This mechanic ensures accessibility



XSL•FO
RenderX

for new parents, allowing them to engage with the game without added pressure or stress. Casual mechanics make the game easier to pick up and play whenever parents have a few free moments.

### *Educational Micromessaging*

Integrate educational content subtly throughout the game to provide important information about PND, such as recognizing symptoms and self-care tips. Delivering these messages in a digestible format helps fill knowledge gaps without overwhelming the player.

### *Character Interactions and Empathy-Building*

Allow players to interact with characters representing supportive individuals, such as partners, family members, or health care professionals. These interactions help model healthy communication and build empathy, both for oneself and others.

### *Gamification Elements*

Add elements such as badges, progress tracking, and daily challenges to encourage ongoing engagement. Gamification makes the experience more rewarding and helps players stay motivated, which is especially important for building positive habits.

### *Celebrating Small Wins*

Highlight small victories in the game, such as successfully reaching out for support or completing a relaxation exercise. Positive reinforcement builds resilience and emphasizes that progress, no matter how small, is meaningful.

## Strengths and Limitations

This review draws strength from its interdisciplinary synthesis of literature spanning digital mental health interventions and serious games, offering a comprehensive overview that bridges technological and psychological perspectives. Examining a novel and emerging approach to PND care contributes to a growing area of interest that remains underexplored in current research. In instances where PND-specific evidence is limited, the review draws from broader depression literature to extract relevant insights while maintaining awareness of contextual differences. Importantly, it also provides a forward-looking perspective by identifying clear gaps in the existing literature and outlining specific directions for future research. These include the development and tailoring of serious games for PND populations, informed by existing evidence on effective game mechanics in comparable mental health contexts.

However, this review also has several limitations. First, the limited number of studies specifically focused on PND constrained the depth of analysis and made it necessary to draw from broader depression literature. While this allowed for the extraction of relevant insights, it may affect the generalizability of findings to the unique psychological and social experience of individuals with PND. Second, there is a potential for selection bias in the inclusion of literature; however, attempts to reduce this included using a structured search process and adhering to the preagreed scope. Future reviews might consider incorporating elements of systematic review methodology, such as dual screening or formal bias appraisal, to enhance rigor. Third, the review was limited to English-language publications, which may have excluded relevant studies from culturally and linguistically diverse populations, potentially narrowing the applicability of findings across global contexts. Expanding the language scope in future reviews could provide a more inclusive understanding of how serious games may support help-seeking across varied cultural settings. Finally, while a narrative review design can lack a systematic or quantitative appraisal of study quality and findings, in this case, given the heterogeneity of the studies reviewed, it did allow for a comprehensive synthesis of the available literature to better pinpoint specific targeted areas needing additional research with recommendations for targeted and novel game design to address that identified gap.

## Future Directions

Based on the findings of this review, the authors plan to develop a game design prototype that selectively incorporates these recommended game mechanics, which will allow for an empirical investigation of utility and efficacy. The game design will have a particular focus on the accessibility and usability needs of new parents. We will begin by conducting user research, including interviews with new parents, to further understand their challenges and preferences. This process will ensure that the game design aligns well with their experiences and provides meaningful support. We will also collaborate interprofessionally with mental health professionals to ensure that the educational content is accurate and that the overall gameplay promotes mental well-being.

Additionally, we aim to develop a hybrid solution that allows for both mobile and web-based accessibility, ensuring that new parents have multiple ways to engage with the game based on their convenience. Once the prototype is ready, we will conduct usability testing and content value testing to assess its possibility and feasibility in engaging new parents and providing mental health support. Feedback from this testing phase will guide further refinements to make the game as supportive and accessible as possible for new parents to improve help-seeking behavior for PND.

## Conclusions

The evidence base for PND-related interventions is currently limited and emerging, highlighting a notable gap in future research. Based on the analysis of existing games, many of them aim to educate players on mental health topics, using engaging scenarios that require players to make supportive decisions and learn about the lived experiences of people with mental health issues. This approach encourages players to understand mental health, corrects misconceptions, and promotes empathy, ultimately reducing stigma. Narrative storytelling also plays a significant role, allowing players to experience the emotional highs and lows of a character's journey, fostering a deep connection and validation for players facing similar struggles. Empowering players was also found to be valuable in seeking professional help by embedding scenarios that encourage reaching out to health care professionals and practicing coping skills in a low-risk, game-based environment. It is argued that these mechanics help players feel more comfortable about seeking support and build their confidence in managing their own mental health.



XSL•FO
RenderX

The included games target a range of audiences, including adolescents, adults, and the general public, identifying a literature gap that only a small number are specifically designed for new parents. New parents face unique challenges, such as exhaustion and time constraints, suggesting the ideal game platform is likely to be mobile—rather than immersive—platforms as they offer the flexibility and ease of use necessary for new parents. Finally, this review underscores the importance of expanding serious game research to more effectively reach and support those experiencing PND, a population that remains inadequately treated and frequently stigmatized in modern culture.

## Acknowledgments

We thank Grammarly (Superhuman Platform) for assistance with proofreading. We used the generative artificial intelligence tool ChatGPT by OpenAI to help with expanding the search terms in the search process.

## Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this paper.

## Authors' Contributions

Conceptualization: WW (lead), EP (equal), VG (equal), JG (equal)
Data curation: WW
Formal analysis: WW
Investigation: WW
Methodology: WW (lead), EP (equal), VG (equal), JG (equal)
Supervision: JG (lead), EP (equal), VG (supporting)
Writing – original draft: WW
Writing – review & editing: EP (lead), JG (equal)

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed search strategies.
[DOCX File, 16 KB - i-jmr_v15i1e70777_app1.docx ]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 106 KB - i-jmr_v15i1e70777_app2.pdf ]

## References

1. Carlson K, Mughal S, Azhar Y, Siddiqui W. Postpartum depression. In: StatPearls: StatPearls Publishing; 2024. URL: https://www.ncbi.nlm.nih.gov/sites/books/NBK519070/ [accessed 2025-12-16]
2. Perinatal depression. National Institute of Mental Health. URL: https://www.nimh.nih.gov/health/publications/perinatal-depression [accessed 2025-12-31]
3. Putnick DL, Sundaram R, Bell EM, et al. Trajectories of maternal postpartum depressive symptoms. Pediatrics 2020 Nov;146(5):e20200857. [doi: 10.1542/peds.2020-0857] [Medline: 33109744]
4. Slomian J, Honvo G, Emonts P, Reginster JY, Bruyère O. Consequences of maternal postpartum depression: a systematic review of maternal and infant outcomes. Womens Health (Lond) 2019;15:1745506519844044. [doi: 10.1177/1745506519844044] [Medline: 31035856]
5. Llewellyn AM, Stowe ZN, Nemeroff CB. Depression during pregnancy and the puerperium. J Clin Psychiatry 1997;58 Suppl 15(15):26-32. [Medline: 9427874]
6. Field T. Postpartum depression effects on early interactions, parenting, and safety practices: a review. Infant Behav Dev 2010 Feb;33(1):1-6. [doi: 10.1016/j.infbeh.2009.10.005] [Medline: 19962196]
7. ACOG Committee on Clinical Practice Guidelines–Obstetrics. Treatment and management of mental health conditions during pregnancy and postpartum. Obstet Gynecol 2023;141(6):1262-1288. [doi: 10.1097/AOG.0000000000005202]
8. Wang Z, Liu J, Shuai H, et al. Correction: mapping global prevalence of depression among postpartum women. Transl Psychiatry 2021;11(1):543. [doi: 10.1038/s41398-021-01692-1]



XSL•FO
RenderX

9. The cost of PNDA. Gidget Foundation Australia. 2019. URL: https://www.gidgetfoundation.org.au/health-professionals/the-cost-of-pnda [accessed 2025-12-24]
10. Luca DL, Margiotta C, Staatz C, Garlow E, Christensen A, Zivin K. Financial toll of untreated perinatal mood and anxiety disorders among 2017 births in the United States. Am J Public Health 2020 Jun;110(6):888-896. [doi: 10.2105/AJPH.2020.305619] [Medline: 32298167]
11. Bauer A, Parsonage M, Knapp M, Iemmi V, Adelaja B, Hogg S. The Costs of Perinatal Mental Health Problems: London School of Economics and Political Science; 2021. URL: https://eprints.lse.ac.uk/59885/1/__lse.ac.uk_storage_LIBRARY_Secondary_libfile_shared_repository_Content_Bauer%2C%20M_Bauer_Costs_perinatal_%20mental_2014_Bauer_Costs_perinatal_mental_2014_author.pdf [accessed 2025-12-16]
12. Fonseca A, Gorayeb R, Canavarro MC. Women s help-seeking behaviours for depressive symptoms during the perinatal period: socio-demographic and clinical correlates and perceived barriers to seeking professional help. Midwifery 2015 Dec;31(12):1177-1185. [doi: 10.1016/j.midw.2015.09.002] [Medline: 26433622]
13. Grissette BG, Spratling R, Aycock DM. Barriers to help-seeking behavior among women with postpartum depression. J Obstet, Gynecol Neonat Nurs 2018 Nov;47(6):812-819. [doi: 10.1016/j.jogn.2018.09.006]
14. Mouaheb H, Fahli A, Moussetad M, Eljamali S. The serious game: what educational benefits? Procedia - Soc Behav Sci 2012;46:5502-5508. [doi: 10.1016/j.sbspro.2012.06.465]
15. Azadegan A, Riedel JC, Baalsrud Hauge J. Serious games adoption in corporate training. Presented at: Serious Games Development and Applications: Third International Conference, SGDA 2012; Sep 26-29, 2012; Bremen, Germany p. 74-85. [doi: 10.1007/978-3-642-33687-4_6]
16. Higham H. Simulation past, present and future-a decade of progress in simulation-based education in the UK. BMJ Simul Technol Enhanc Learn 2021;7(5):404-409. [doi: 10.1136/bmjstel-2020-000601] [Medline: 35515719]
17. Schöbel SM, Janson A, Söllner M. Capturing the complexity of gamification elements: a holistic approach for analysing existing and deriving novel gamification designs. Eur J Inf Syst 2020 Nov 1;29(6):641-668. [doi: 10.1080/0960085X.2020.1796531]
18. Ruiz M, Moreno M, Girela-Serrano B, et al. Winning the game against depression: a systematic review of video games for the treatment of depressive disorders. Curr Psychiatry Rep 2022 Jan;24(1):23-35. [doi: 10.1007/s11920-022-01314-7]
19. Townsend C, Humpston C, Rogers J, Goodyear V, Lavis A, Michail M. The effectiveness of gaming interventions for depression and anxiety in young people: systematic review and meta-analysis. BJPsych Open 2022 Jan 7;8(1):e25. [doi: 10.1192/bjo.2021.1078] [Medline: 34991767]
20. Solinska-Nowak A, Magnuszewski P, Curl M, et al. An overview of serious games for disaster risk management – prospects and limitations for informing actions to arrest increasing risk. Int J Disaster Risk Reduct 2018 Oct;31:1013-1029. [doi: 10.1016/j.ijdrr.2018.09.001]
21. van den Heuvel JF, Groenhof TK, Veerbeek JH, et al. eHealth as the next-generation perinatal care: an overview of the literature. J Med Internet Res 2018;20(6):e202. [doi: 10.2196/jmir.9262]
22. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016 Dec;16(1):1-10. [doi: 10.1186/s12884-016-0856-5]
23. O'Brien E, Rauf Z, Alfirevic Z, Lavender T. Women's experiences of outpatient induction of labour with remote continuous monitoring. Midwifery 2013 Apr;29(4):325-331. [doi: 10.1016/j.midw.2012.01.014] [Medline: 23159160]
24. Ming WK, Mackillop LH, Farmer AJ, et al. Telemedicine technologies for diabetes in pregnancy: a systematic review and meta-analysis. J Med Internet Res 2016 Nov 9;18(11):e290. [doi: 10.2196/jmir.6556] [Medline: 27829574]
25. Lau Y, Htun TP, Wong SN, Tam WSW, Klainin-Yobas P. Therapist-supported internet-based cognitive behavior therapy for stress, anxiety, and depressive symptoms among postpartum women: a systematic review and meta-analysis. J Med Internet Res 2017 Apr 28;19(4):e138. [doi: 10.2196/jmir.6712] [Medline: 28455276]
26. Sondaal SFV, Browne JL, Amoakoh-Coleman M, et al. Assessing the effect of mhealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. PLOS ONE 2016;11(5):e0154664. [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
27. Tapia-Conyer R, Lyford S, Saucedo R, et al. Improving perinatal care in the rural regions worldwide by wireless enabled antepartum fetal monitoring: a demonstration project. Int J Telemed Appl 2015;2015(1):794180. [doi: 10.1155/2015/794180] [Medline: 25691900]
28. Hussain-Shamsy N, Shah A, Vigod SN, Zaheer J, Seto E. Mobile health for perinatal depression and anxiety: scoping review. J Med Internet Res 2020 Apr 13;22(4):e17011. [doi: 10.2196/17011] [Medline: 32281939]
29. Dosani A, Arora H, Mazmudar S. mHealth and perinatal depression in low-and middle-income countries: a scoping review of the literature. IJERPH 2020;17(20):7679. [doi: 10.3390/ijerph17207679]
30. Wan Mohd Yunus WMA, Matinolli HM, Waris O, et al. Digitalized cognitive behavioral interventions for depressive symptoms during pregnancy: systematic review. J Med Internet Res 2022 Feb 23;24(2):e33337. [doi: 10.2196/33337] [Medline: 35195532]
31. Lau Y, Yen KY, Wong SH, Cheng JY, Cheng LJ. Effect of digital cognitive behavioral therapy on psychological symptoms among perinatal women in high income-countries: a systematic review and meta-regression. J Psychiatr Res 2022 Feb;146:234-248. [doi: 10.1016/j.jpsychires.2021.11.012] [Medline: 34774299]




XSL•FO
RenderX

32. Fallon V, Davies SM, Silverio S, Creagh L. Virtual reality interventions designed to support parents during and throughout the first year after birth: a scoping review. Digit Health 2024;10:20552076241245373. [doi: 10.1177/20552076241245373] [Medline: 38655377]
33. Motrico E, Conejo-Cerón S, Martín-Gómez C, Gómez-Gómez I, Fonseca A, Moreno-Peral P. Effectiveness of web-based and mobile-based psychological interventions to prevent perinatal depression: study protocol for a systematic review and meta-analysis of randomized controlled trials. Internet Interv 2021 Dec;26:100471. [doi: 10.1016/j.invent.2021.100471]
34. Spadaro B, Martin-Key NA, Funnell E, Bahn S. mHealth solutions for perinatal mental health: scoping review and appraisal following the mHealth index and navigation database framework. JMIR mHealth uHealth 2022 Jan 17;10(1):e30724. [doi: 10.2196/30724] [Medline: 35037894]
35. Pilkington PD, Whelan TA, Milne LC. A review of partner‐inclusive interventions for preventing postnatal depression and anxiety. Clin Psychol (Aust Psychol Soc) 2015 Jul 1;19(2):63-75. [doi: 10.1111/cp.12054]
36. Liberto TL. Screening for depression and help-seeking in postpartum women during well-baby pediatric visits: an integrated review. J Pediatr Health Care 2012 Mar;26(2):109-117. [doi: 10.1016/j.pedhc.2010.06.012] [Medline: 22360930]
37. What is perinatal depression? American Psychiatric Association. URL: https://www.psychiatry.org/patients-families/peripartum-depression/what-is-peripartum-depression [accessed 2026-01-15]
38. Fleming T, Cheek C, Merry SN, et al. Serious games for the treatment or prevention of depression: a systematic review. Rev Psicopatologia Psicologia Clin 2014;19(3):227-242. [doi: 10.5944/rppc.vol.19.num.3.2014.13904]
39. Li J, Theng YL, Foo S. Game-based digital interventions for depression therapy: a systematic review and meta-analysis. Cyberpsychol Behav Soc Netw 2014 Aug;17(8):519-527. [doi: 10.1089/cyber.2013.0481] [Medline: 24810933]
40. Dias LPS, Barbosa JLV, Vianna HD. Gamification and serious games in depression care: a systematic mapping study. Telematics Inf 2018 Apr;35(1):213-224. [doi: 10.1016/j.tele.2017.11.002]
41. Kowal M, Conroy E, Ramsbottom N, Smithies T, Toth A, Campbell M. Gaming your mental health: a narrative review on mitigating symptoms of depression and anxiety using commercial video games. JMIR Serious Games 2021 Jun 16;9(2):e26575. [doi: 10.2196/26575] [Medline: 34132648]
42. Pine R, Fleming T, McCallum S, Sutcliffe K. The effects of casual videogames on anxiety, depression, stress, and low mood: a systematic review. Games Health J 2020 Aug;9(4):255-264. [doi: 10.1089/g4h.2019.0132] [Medline: 32053021]
43. King M, Marsh T, Akcay Z. A review of indie games for serious mental health game design. Presented at: Serious Games: Joint International Conference, JCS; Jan 12-13, 2022; Virtual Event p. 138-152. [doi: 10.1007/978-3-030-88272-3_11]
44. Martinez K, Menéndez-Menéndez MI, Bustillo A. Awareness, prevention, detection, and therapy applications for depression and anxiety in serious games for children and adolescents: systematic review. JMIR Serious Games 2021;9(4):e30482. [doi: 10.2196/30482]
45. Yen HY, Chiu HL. Virtual reality exergames for improving older adults' cognition and depression: a systematic review and meta-analysis of randomized control trials. J Am Med Dir Assoc 2021 May;22(5):995-1002. [doi: 10.1016/j.jamda.2021.03.009] [Medline: 33812843]
46. Abd-Alrazaq A, Al-Jafar E, Alajlani M, et al. The effectiveness of serious games for alleviating depression: systematic review and meta-analysis. JMIR Serious Games 2022;10(1):e32331. [doi: 10.2196/32331]
47. Kim Y, Hong S, Choi M. Effects of serious games on depression in older adults: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022;24(9):e37753. [doi: 10.2196/37753]
48. Chitale V, Baghaei N, Playne D, et al. The use of videogames and virtual reality for the assessment of anxiety and depression: a scoping review. Games Health J 2022 Jul 26(6):341-354. [doi: 10.1089/g4h.2021.0227] [Medline: 35881854]
49. Gliosci R, Silva TBPE. Therapeutic interventions with videogames in treatments for depression: a systematic review. Games Health J 2023 Aug;12(4):269-279. [doi: 10.1089/g4h.2022.0094] [Medline: 36946750]
50. Gómez-León MI. Serious games to support emotional regulation strategies in educational intervention programs with children and adolescents. Systematic review and meta-analysis. Heliyon 2025 Feb;11(4):e42712. [doi: 10.1016/j.heliyon.2025.e42712]
51. First MB. Diagnostic and Statistical Manual of Mental Disorders, 5th edition, and clinical utility. J Nerv Ment Dis 2013 Sep;201(9):727-729. [doi: 10.1097/NMD.0b013e3182a2168a] [Medline: 23995026]
52. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE Guide No. 131. Med Teach 2020 Aug;42(8):846-854. [doi: 10.1080/0142159X.2020.1755030] [Medline: 32356468]
53. Naeem M, Ozuem W, Howell K, Ranfagni S. A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. Int J Qual Methods 2023 Oct;22. [doi: 10.1177/16094069231205789]
54. Wehbe RR, Watson DK, Tondello GF, et al. Above Water: an educational game for anxiety. 2016 Oct Presented at: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts; Oct 16-19, 2016 p. 79-84. [doi: 10.1145/2968120.2971804]
55. Tuijnman A, Kleinjan M, Hoogendoorn E, Granic I, Engels RC. A game-based school program for mental health literacy and stigma regarding depression (Moving Stories): protocol for a randomized controlled trial. JMIR Res Protoc 2019 Mar 14;8(3):e11255. [doi: 10.2196/11255] [Medline: 30869652]




XSL•FO
RenderX

56. Cheng VWS, Davenport TA, Johnson D, Vella K, Mitchell J, Hickie IB. An app that incorporates gamification, mini-games, and social connection to improve men's mental health and well-being (MindMax): participatory design process. JMIR Ment Health 2018 Nov 19;5(4):e11068. [doi: 10.2196/11068] [Medline: 30455165]
57. Mueller S, Soriano D, Boscor A, et al. MANTRA: development and localization of a mobile educational health game targeting low literacy players in low and middle income countries. BMC Public Health 2020 Jul 28;20(1):1171. [doi: 10.1186/s12889-020-09246-8] [Medline: 32723317]
58. Sara TA, Nazifa SN, Tasneem S, Shakib TU, Islam MN. PPDHero: requirements elicitation and development of a system to empower new mothers on postpartum depression. 2024 Aug Presented at: International Conference on Intelligent Systems Design and Applications; Dec 11-13, 2023 p. 389-398. [doi: 10.1007/978-3-031-64850-2_37]
59. Cangas AJ, Navarro N, Parra JMA, et al. Stigma-Stop: a serious game against the stigma toward mental health in educational settings. Front Psychol 2017;8:1385. [doi: 10.3389/fpsyg.2017.01385] [Medline: 28878702]
60. Hanisch SE, Birner UW, Oberhauser C, Nowak D, Sabariego C. Development and evaluation of digital game-based training for managers to promote employee mental health and reduce mental illness stigma at work: quasi-experimental study of program effectiveness. JMIR Ment Health 2017 Aug 4;4(3):e31. [doi: 10.2196/mental.7600] [Medline: 28778839]
61. Ferchaud A, Seibert J, Sellers N, Escobar Salazar N. Reducing mental health stigma through identification with video game avatars with mental illness. Front Psychol 2020;11:2240. [doi: 10.3389/fpsyg.2020.02240] [Medline: 33013585]
62. Depression Quest. URL: http://www.depressionquest.com/ [accessed 2025-12-16]
63. Keep in Mind: Remastered. Steam. URL: https://store.steampowered.com/app/643200 [accessed 2025-12-16]
64. Before I Forget. Steam. URL: https://store.steampowered.com/app/1126600/Before [accessed 2025-12-16]
65. Sea of Solitude. Wikipedia. URL: https://en.wikipedia.org/wiki/Sea_of_Solitude [accessed 2025-12-16]
66. Rodrigues MAF, de Oliveira TRC, de Figueiredo DL, et al. An interactive story decision-making game for mental health awareness. 2022 Aug Presented at: 2022 IEEE 10th International Conference on Serious Games and Applications for Health(SeGAH); Aug 10-12, 2022; Sydney, Australia p. 1-8. [doi: 10.1109/SEGAH54908.2022.9978592]
67. Taubman-Ben-Ari O, Ben-Yaakov O, Chasson M. Parenting stress among new parents before and during the COVID-19 pandemic. Child Abuse Neglect 2021 Jul;117:105080. [doi: 10.1016/j.chiabu.2021.105080] [Medline: 33930664]
68. Saxbe D, Rossin-Slater M, Goldenberg D. The transition to parenthood as a critical window for adult health. Am Psychol 2018 Dec;73(9):1190-1200. [doi: 10.1037/amp0000376] [Medline: 30525801]
69. Msetfi RM, Kornbrot DE, Halbrook YJ. The association between the sense of control and depression during the COVID-19 pandemic: a systematic review and meta-analysis. Front Psychiatry 2024;15:1323306. [doi: 10.3389/fpsyt.2024.1323306] [Medline: 38414499]
70. Liu Y, Yang Y, Bai X, Chen Y, Mo L. Do immediate external rewards really enhance intrinsic motivation? Front Psychol 2022;13:853879. [doi: 10.3389/fpsyg.2022.853879]
71. Moore RL, Hwang W, Moses JD. A systematic review of mobile-based microlearning in adult learner contexts. Educ Technol Soc 2024;27(1):137-146. [doi: 10.30191/ETS.202401_27(1).SP02]


## Abbreviations

**CBT:** cognitive behavioral therapy
**CVG:** casual video game
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition
**mHealth:** mobile health
**PND:** postnatal depression
**VR:** virtual reality


*Edited by N Cahill; submitted 01.Jan.2025; peer-reviewed by A Gazis, M Evans; revised version received 01.Dec.2025; accepted 01.Dec.2025; published 16.Jan.2026.*

*Please cite as:*
*Wang W, Penney E, Gay V, Garcia J*
*Exploring Serious Games in Supporting Postnatal Depression: Narrative Review*
*Interact J Med Res 2026;15:e70777*
*URL: https://www.i-jmr.org/2026/1/e70777*
*doi:10.2196/70777*


© Weijie Wang, Erika Penney, Valerie Gay, Jaime Garcia. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 16.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction



XSL•FO
RenderX

in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Identifying Key Predictors of Appropriate Discharge Destinations for Older Inpatients in Acute Care: Scoping Review

Christoph Leinert[1], Dr med; Marina Liselotte Fotteler[1,2], Dr biol hum; Thomas Derya Kocar[1], Dr med; Jessica Wolf[1]; Lisa Beissel[1], BSc; Kathrin Grummich[3], Diplom (FH); Dhayana Dallmeier[1,4], Prof Dr med, PhD; Michael Denkinger[1], Prof Dr med

[1]Institute for Geriatric Research at Agaplesion Bethesda Ulm, Ulm University Medical Center, Zollernring 26, Ulm, Germany

[2]DigiHealth Institute, Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany

[3]ZB MED - Information Centre for Life Sciences, Cologne, Germany

[4]Department of Epidemiology, Boston University School of Public Health, Boston, MA, United States

**Corresponding Author:**
Christoph Leinert, Dr med
Institute for Geriatric Research at Agaplesion Bethesda Ulm, Ulm University Medical Center, Zollernring 26, Ulm, Germany

## Abstract

**Background:** Postacute care (PAC) services are important to ensure functional recovery and provide adequate care for geriatric inpatients in acute care. The choice between different PAC options can be challenging, and predictors for the most appropriate among diverse discharge options are warranted.

**Objective:** We conducted a scoping review to identify predictors of appropriate discharge destinations for older adults (≥65 y) in acute care transitioning to different PAC settings and extract the most relevant predictors for different PAC settings as well as a generalizable set of predictor domains.

**Methods:** The databases of Medline, Embase, Cochrane Central Register of Controlled Trials, PsycINFO, CINAHL, and Emcare were systematically searched for English or German literature published until February 25, 2022. A total of 3 researchers screened, extracted, and categorized the data according to domains, discharge destinations, mean age, and health care systems origin, focusing on predictors that increase the likelihood of a discharge destination (positive predictors). The Jaccard index was calculated to compare the similarity between different possible domain combinations and existing literature.

**Results:** Of 22,382 records screened, 171 quantitative and 10 qualitative studies were included. After separating combined discharge destinations, we found 1047 predictors for different discharge destinations including nursing home (n=297, 28%), skilled nursing facility (n=223, 21%), inpatient rehabilitation (n=206, 20%), home with (n=97, 9%) or without (n=74, 7%) support, assisted living (n=63, 6%), and early inpatient rehabilitation (n=21, 2%). Of all positive predictors (n=723), age was the most frequently reported predictor (80/723, 11%). Geriatric syndromes were found more often in patients 80 years or older (121/192, 63%) and in non-US studies (174/285, 61%). The top reported predictors for discharge to nursing homes were diagnosed dementia (9/297, 3%) and deficits in instrumental activities of daily living (ADL; 10/297, 3%); for discharge to rehabilitation, the top predictors were longer length of stay (11/205, 5%) and existent cardiopulmonary disease (10/205, 5%); and for back home without support, the top predictors were good ADL (10/74, 14%) and mobility assessments (9/74, 12%). Among 20 predictor domains, 8 were most concordant with the literature: cognitive impairment, ADL, demographics, social support, hospitalization data, multimorbidity, mobility, and primary diagnosis.

**Conclusions:** This scoping review provides a comprehensive overview of predictors for appropriate discharge decisions in older adults in acute care, stratified by destination, age, study origin, and the predictor domains most concordant with the literature. The results will be valuable to inform the choice of features for clinical decision support systems, including the training of machine learning algorithms.

***(Interact J Med Res 2026;15:e76582)*** doi:10.2196/76582

**KEYWORDS**

CDSS; clinical decision support system; continuity of care; discharge destination; discharge planning; feature selection; geriatrics; postacute care; predictor



XSL•FO
RenderX

## Introduction

Geriatric inpatients are at a high risk of functional decline during acute care treatment, and post-acute care (PAC) services are needed to ensure recovery [1]. Effective discharge planning has been shown to mitigate readmission rates, reduce hospital stays, lower associated costs, and enhance patient satisfaction [2]. Thus, choosing the most appropriate discharge destination is vital within geriatric comanagement, guaranteeing alignment with individualized rehabilitation and care needs. Modern health care systems present diverse PAC options, making selection challenging and highlighting the need for predictors of appropriate discharge destinations [3-6].

The World Health Organization defines continuity of care (COC) as the degree to which health care events are perceived as connected and coherent over time, aligning with patients' health necessities and preferences [7]. COC can be further categorized into relational continuity (patient-provider relationship), informational continuity (communication), and management continuity (coordination) [8]. In the context of geriatric comanagement in acute inpatient care, the choice of the most appropriate discharge destination refers to the management continuity aspect. In addition to COC and discharge planning, various terms encompass this dimension, including integrated care, case management, or transitional care [9,10].

Existing research has focused on predictors differentiating binary outcomes [11-13]. However, there remains a need for predictors supporting decisions among multiple discharge options to not miss relevant features.

To address this gap, we conducted a scoping review to identify potential predictors for the most appropriate discharge destination for older inpatients in acute care transitioning to different PAC options such as outpatient, inpatient or early rehabilitation, skilled nursing facility (SNF), nursing home, assisted living, or home-based care with or without support. As different health care systems offer different discharge destinations and funding options, we also planned to stratify according to the different health care system origins. Our scoping review will inform the feature selection for the development of a machine learning−driven clinical decision support system (CDSS) within the "Supporting Surgery with Geriatric Co-Management and AI" (SURGE-Ahead) project by providing a broad overview of predictive measures for different discharge options [14].

## Methods

### Conceptualization

We conducted a scoping review adhering to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Scoping Review extension guidelines; Checklist 1) and the Joanna Briggs Institute guidance [15,16]. A review protocol was registered on Open Science Framework and was adapted after a piloting phase to focus on "predictors" instead of "predictors and outcome measures" due to a lack of specific outcome measures for COC identified [17].

### Eligibility Criteria

Using the Joanna Briggs Institute's population-concept-context framework, our focus was on older adult inpatients 65 years and older treated in acute surgical or medical hospital departments [18]. We included only participants 65 years and older, verifying this by examining the inclusion criteria and the reported age structure of each study population. Studies were included if they reported an age range with a lower limit of 65 years or above or at least provided a mean or median of 65 years or more. Publications with unclear age demographics were conservatively excluded. We aimed to identify predictors utilized by health care professionals to determine the most appropriate PAC setting. Predictors were defined as discrete, separable parameters associated with a specific discharge destination. When faced with complex models combining multiple parameters, we included the individual parameters whenever feasible but not the complex model itself. In quantitative analyses, we further required evidence of a statistically significant effect ($P \leq .05$). Studies had to involve transitions from acute care to different PAC options. German- or English-language publications were considered, excluding letters, comments, case studies, editorials, and studies primarily addressing health economic issues or involving psychiatry or rehabilitation departments for older adults. We excluded studies that chose to report composite discharge options like "discharge to nursing home or death," as these outcomes should not be combined. Similarly, studies including palliative care services like hospice as part of a composite discharge destination were excluded, as we consider individualized decision-making essential for palliative care.

### Search Strategy

Comprehensive search strategies were developed that included different concepts for the successful choice of discharge destinations, including continuity of care, coordination of care, transition of care, integrated health care, case management, discharge planning, or rehabilitation eligibility determination. On February 25, 2022, the databases Medline (OVID interface), Embase (OVID), Cochrane Central Register of Controlled Trials (Wiley), PsycINFO (EBSCOhost), CINAHL (EBSCOhost), and Emcare (OVID) were systematically searched, guided by an experienced information specialist (KG). Search strategies for all databases searched are shown in Multimedia Appendix 1.

### Study Selection and Data Extraction

Records obtained through the database search were imported into the Covidence systematic review software (Veritas Health Innovation). Additionally, the primary publications of all reviews identified in the literature search were imported into Covidence. Duplicates were removed using the deduplication functionality of the software, followed by manual verification in unclear cases marked by the software (MLF and CL). Two out of the 4 reviewers (MLF, CL, LB, and JW) independently screened all titles and abstracts for eligibility, followed by full-text screening. Any discrepancies were resolved through adjudication by a third author. Data extraction for each article was conducted by 2 of the 3 authors independently (MLF, CL, and JW) and consolidated by the team. The extracted data



XSL•FO
RenderX

encompassed study and population characteristics, health care systems origin, discharge destination, statistical analysis type, predictor, and quantitative measures of predictive strength (odds ratio, relative risk, confidence interval, *P* value) if available. The identified predictors were categorized into different domains after discussion in the review team. Data extraction for qualitative studies was conducted separately using a thematic evidence synthesis, assigning all qualitative and quantitative predictors, including representative quotes, to the predefined domains [19].

### Data Analysis

Predictive directions indicating an increased (positive) or reduced (negative) probability of discharge to a specific environment were determined from odds ratios, relative risks, or manual labeling and joint consensus when necessary (MLF and CL). The negative predictors reported in the included studies often used double negations or lacked a differentiation of discharge alternatives. To avoid misdirected results, we focused our analyses on the positive predictors.

Predictors were also stratified based on discharge destinations (outpatient, inpatient or early rehabilitation, SNF, nursing home, assisted living, or home-based care with or without support, long-term acute care setting, or other acute care setting). In studies that consolidated multiple discharge destinations into a single outcome, these were extracted separately. The direction of effect was added to the top 5 predictors per discharge destination via manual extraction from the literature and joint consensus labeling (MLF and CL). Additionally, we stratified by mean population age (<80 y or ≥80 y) and health care systems origin: predictors from Anglo-European studies with mostly publicly funded health care systems (Europe, Canada, Australia) were also analyzed separately from studies from the United States of America with a relevant proportion of pay-for-service health services [20-22].

Similar predictive factors identified across multiple studies were consolidated, whereas validated assessment tools were treated independently. In some studies, data on predictive measures were reported without using a validated assessment instrument. In these cases, we categorized the results as “no specific score.” As an example, in the activities of daily living (ADL) domain, many studies reported “decreased ADL” or similar but did not use an established assessment like the Barthel index (BI). In these cases, we categorized the reported predictors as “ADL no specific score.”

The Jaccard Index can be used to measure similarity and diversity of sample sets [23]. It was used to determine the concordance between different potential domain combinations and those reported in the reviewed studies. A higher index value signifies better agreement. Based on the 20 defined domains, we juxtaposed approximately 1 million potential combination sets (all possible combinations) to be compared with those found in the literature. The domain set with the highest average Jaccard index value was identified as a potentially generalizable combination of key predictor domains most commonly associated with each other and existing literature.

Descriptive statistics and data manipulation were carried out using Microsoft Excel 2019 (Microsoft). We used Python 3.12.2 for the analysis, using the itertools standard library for generating combinations and set class methods for calculating the Jaccard index. For loading the data from Excel, the pandas 2.2.1 library was used.

## Results

### Overview

Our search yielded 22,382 database entries after removing duplicates. After title or abstract screening, we assessed the full texts of 475 studies for eligibility and included 181 studies (n=171 quantitative studies, n=10 qualitative studies). Figure 1 shows the PRISMA flow chart of the scoping review. All included studies are listed in Multimedia Appendix 2.

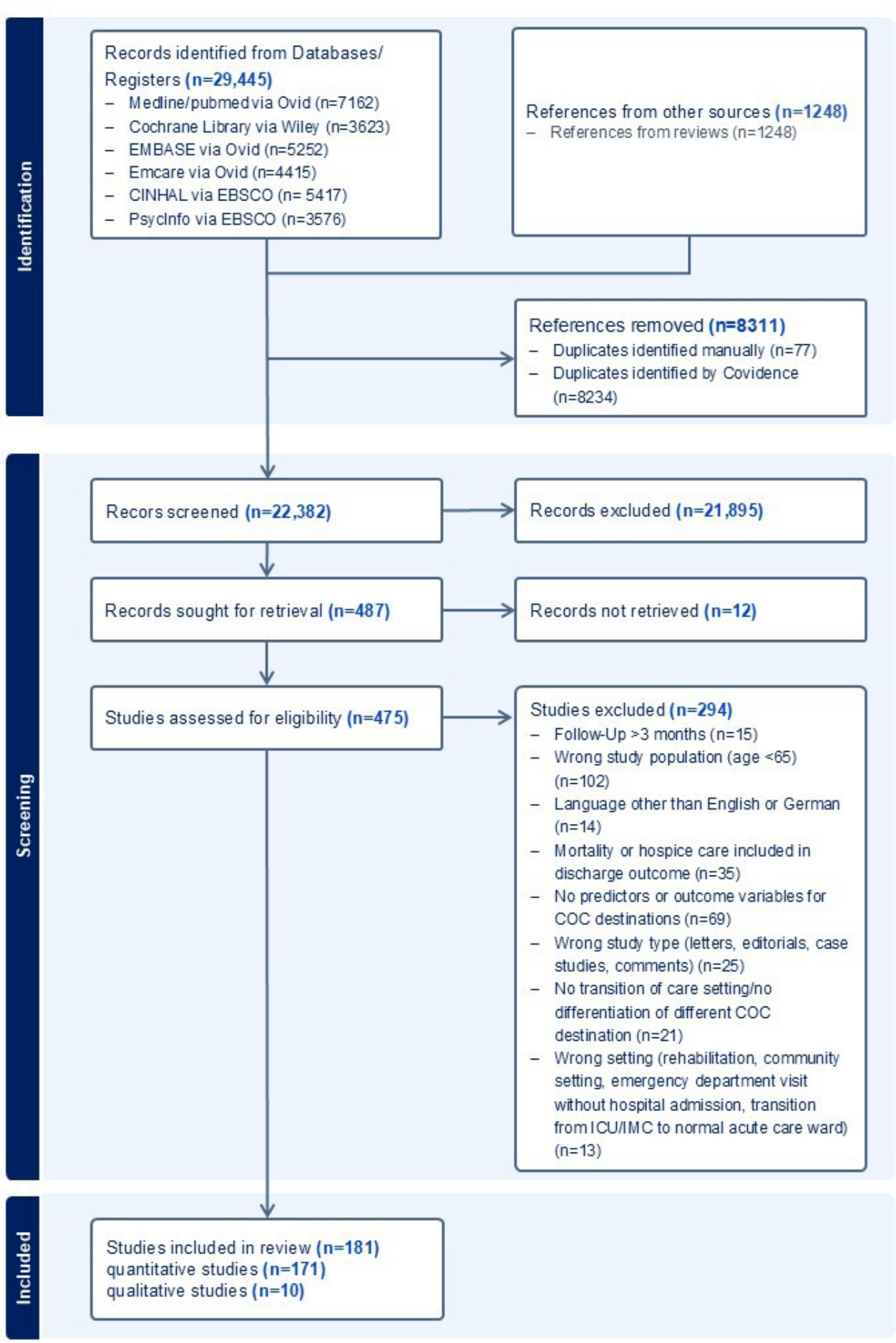


XSL•FO
RenderX

**Figure 1.** PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flowchart. COC: continuity of care; ICU: intensive care unit; IMC: intermediate care unit.



XSL•FO
RenderX






XSL•FO
RenderX

The included studies were published between 1984 and 2022. Most of the 181 studies originated from North America (USA: n=84, 46%; Canada: n=14, 8%). Europe accounted for approximately one-third (n=60, 33%), including studies from the United Kingdom (n=11, 6%), the Netherlands (n=11, 6%), Switzerland (n=8, 4%), Germany (n=7, 4%), Italy (n=5, 3%), and Spain (n=3, 2%). The remaining 15% were conducted in other parts of the world, including Australia (n=14, 8%) and Japan (n=7, 4%).

A total of 275 (in part combined) settings were described, and most studies (n=161, 58.5%) were conducted in surgical settings such as trauma or orthogeriatrics (n=75, 27%) and general or visceral surgery (n=28, 10%). Other prevalent settings were geriatrics (n=29, 11%) and general or internal medicine (n=27, 10%). In 7% (n=19) of settings, no clinical department was specified. The included studies encompass data from 6,357,026 participants with a mean age of 77.42 years.

### Quantitative Studies

In the 171 quantitative studies, a total of 856 predictors for various discharge destinations were identified. More than half of these studies were prospective (88/171, 51.5%), while the remaining utilized a retrospective (78/171, 46%) or mixed approach (5/171, 3%). Most reported predictors were positive predictors (723/856, 84.5%). About two-thirds of the studies (116/171, 68%) reported the participants' mean age, with 55 of 171 (32%) studies having a mean age of 80 years or older and 61 (36%) having a mean age younger than 80 years.

### Predictors

Age emerged as the most prevalent predictor across all categories. The 10 most frequent positive predictors are listed in Table 1 for all studies, and stratified by mean age <80 years, ≥80 years, US and Anglo-European studies.



XSL•FO
RenderX

**Table .** Top 10 positive predictors for continuity of care of all the included studies, stratified by mean age (<80 vs ≥80 y) and US and Anglo-European origin[a].

| Predictor | All (n=723), n (%) | <80 years mean age (n=274), n (%) | ≥80 years mean age (n=192), n (%) | US (n=413), n (%) | Anglo-European (n=285), n (%) |
|---|---|---|---|---|---|
| Age | 80 (11) | 34 (12) | 20 (10) | 50 (12) | 28 (10) |
| Wound problems | 25 (4) | 23 (8) | —[b] | 25 (6) | — |
| Cardiopulmonary disease | 24 (3) | 16 (6) | — | 23 (6) | — |
| ADL[c] NSS[d] | 23 (3) | 10 (4) | — | 19 (5) | — |
| Length of stay | 22 (3) | 6 (2) | — | — | 14 (5) |
| Infectious disease | 17 (2) | 16 (6) | — | 16 (4) | — |
| Number of comorbidities | 16 (2) | 11 (4) | — | 15 (4) | — |
| ASA[e] score [24] (multimorbidity) | 16 (2) | 9 (3) | — | 13 (3) | — |
| Falls | 14 (2) | — | 5 (3) | 9 (2) | 7 (3) |
| Frailty index[f] | 14 (2) | 8 (3) | — | — | — |
| Female sex | 14 (2) | — | 6 (3) | 9 (2) | — |
| Mobility NSS | — | 7 (3) | 5 (3) | 9 (2) | — |
| Caregiver support | — | 6 (2) | — | — | — |
| Deep vein thrombosis diagnosis | — | 6 (2) | — | — | — |
| Source of admission | — | 6 (2) | — | — | — |
| IADL[g] NSS | — | — | 9 (5) | — | 9 (3) |
| Problems personal hygiene | — | — | 9 (5) | — | 7 (3) |
| Katz score [25] (ADL) | — | — | 8 (4) | — | 11 (4) |
| Charlson comorbidity index [26] | — | — | 7 (4) | — | — |
| Dementia diagnosis | — | — | 6 (3) | — | 7 (3) |
| Cognitive deficits NSS | — | — | 5 (3) | — | — |
| Living with or without companion | — | — | 5 (3) | — | 7 (3) |
| Barthel index [27] (ADL) | — | — | — | — | 7 (3) |
| Injury severity score [28] | — | — | — | — | 7 (3) |

[a]n: number of extracted positive predictors, % proportion of extracted positive predictors in each group.

[b]Not applicable.

[c]ADL: activities of daily living.

[d]NSS: no specific score.

[e]ASA: American Society of Anesthesiologists.

[f]Frailty index: different assessments possible.

[g]IADL: instrumental activities of daily living.

## Discharge Destinations

We analyzed predictor-destination sets (n=1047) across 10 different discharge options. A discharge to nursing homes emerged as the most common option (n=297, 28%), followed by SNF (n=223, 21%), inpatient rehabilitation (n=206, 20%), discharge home with support (n=97, 9%) and without support (n=74, 7%), and assisted living (n=63, 6%). Less frequently

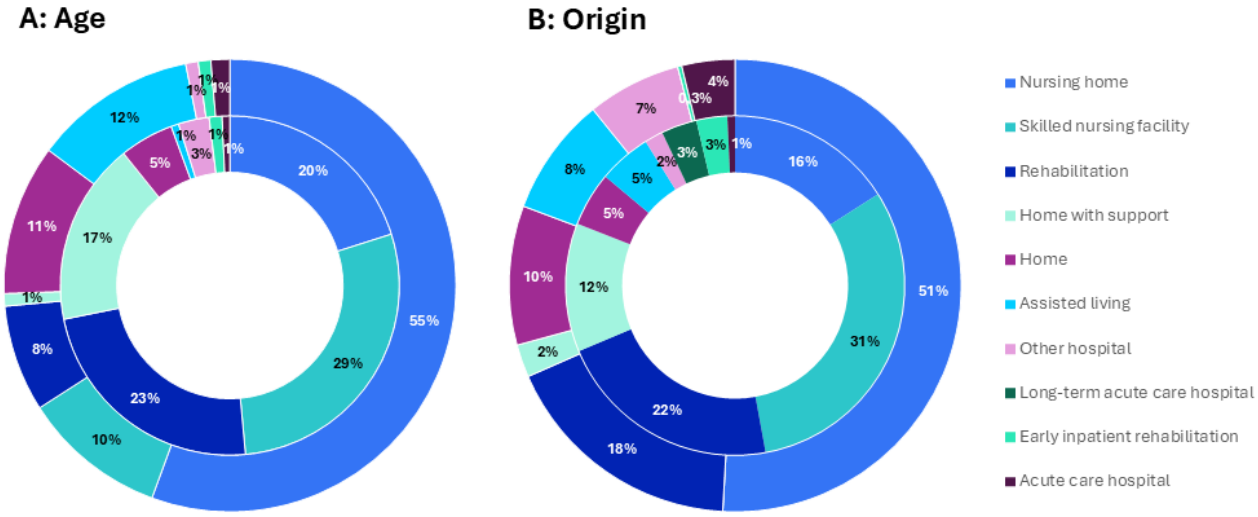


XSL•FO
RenderX

found were predictors related to early inpatient geriatric rehabilitation (n=21, 2%), long-term acute care hospitals (n=23, 2%), and other acute care hospitals (n=18, 2%).

Nursing homes were the predominant destination in the oldest old category (≥80 y: 124/223, 55% vs <80 y: 82/408, 20%), and inpatient rehabilitation (<80 y: 95/408, 23% vs. ≥80 y: 17/223, 8%) and SNF (<80 y: 116/408, 29% vs ≥80 y: 23/223, 10%) were the main destination in the younger old category (see Figure 2A). In Anglo-European studies, the most frequently reported discharge destination was nursing home (174/342, 51%) compared to SNFs (221/677, 31%) in US studies (see Figure 2B).

**Figure 2.** Predictors stratified by discharge destination. (A) Proportion of discharge destinations of studies stratified with mean age <80 years (inner circle, n=408) and ≥80 years (outer circle, n=223). (B) Proportion of discharge destinations of studies originating from the United States (inner circle, n=677) and Anglo-European countries (outer circle, n=342).




## Predictors for Different Discharge Destinations

Age emerged as the predominant positive predictor for most discharge options (Table 2). Nevertheless, for a discharge home, the BI was found more often. Because of the scarcity of predictors in the categories early inpatient rehabilitation, acute care hospital, long-term acute care hospital, and other hospitals, no further stratified analysis was done for these destinations.



XSL•FO
RenderX

**Table .** Top 5 positive predictors for 6 discharge destinations[a].

| Predictor | Nursing home (n=297), n (%) | Skilled nursing facility (n=211), n (%) | Rehabilitation (n=205), n (%) | Home with support (n=97), n (%) | Home (n=74), n (%) | Assisted living (n=63), n (%) |
|---|---|---|---|---|---|---|
| Age (older)[b] | 33 (11) | 24 (11) | 25 (12) | 10 (10) | —[c] | 4 (6) |
| IADL[d] NSS[e] (dependent on the instrument used) | 10 (3) | — | — | — | — | — |
| Dementia diagnosis (yes) | 9 (3) | — | — | — | — | — |
| Length of stay (longer) | 9 (3) | — | 11 (5) | 4 (4) | — | — |
| Sex (female) | 9 (3) | — | — | — | — | 4 (6) |
| ADL[f] NSS (dependent on the instrument used) | 8 (3) | 8 (4) | 9 (4) | 4 (4) | — | — |
| Cardiopulmonary disease (yes) | — | 11 (5) | 10 (5) | 9 (9) | — | — |
| Wound problem (yes) | — | 11 (5) | 10 (5) | 6 (6) | — | — |
| Number of comorbidities (higher) | — | 8 (4) | — | 4 (4) | — | — |
| Frailty index (frailer) | — | 8 (4) | — | — | — | — |
| Infectious disease (yes) | — | — | — | 6 (6) | — | — |
| Age (younger) | — | — | — | — | 4 (5) | — |
| Barthel index [27] (higher) | — | — | — | — | 6 (8) | — |
| Katz score [25] (ADL) (higher) | — | — | — | — | 4 (5) | — |
| Cumulated ambulation score [29] (higher) | — | — | — | — | 3 (4) | — |
| De Morton Mobility Index [30] (higher) | — | — | — | — | 3 (4) | — |
| Short physical performance battery [31] (higher) | — | — | — | — | 3 (4) | — |
| ICU[g] treatment (yes) | — | — | — | — | — | 4 (6) |
| Older people's QoL[h] questionnaire [32] (low) | — | — | — | — | — | 4 (6) |
| Anesthesia or ICU treatment procedure (yes) | — | — | — | — | — | 3 (5) |

[a]n: number of extracted positive predictors, % proportion of extracted positive predictors in each group.

[b]Direction of effect in parentheses.

[c]Not applicable.

[d]IADL: instrumental activities of daily living.

[e]NSS: no specific score.



XSL•FO
RenderX

[f]ADL: activities of daily living.

[g]ICU: intensive care unit.

[h]QoL: quality of life.

Negative predictors were a minority (133/856, 15.5%) and very heterogeneous, and often seemed not useful for the identification of the most appropriate discharge destinations, such as ethnicity or male sex for discharge to rehabilitation. More relevant predictors were the unavailability of caregivers for discharge to a nursing home, a hip fracture diagnosis, or a higher frailty index, not allowing for discharge home.

A detailed overview of the extracted evidence stratified by mean age, study type, study origin, and discharge destination is shown in Multimedia Appendix 3 and, according to the domains, positive as well as negative predictors, predictive strength indicators, and corresponding literature references in Multimedia Appendix 4.

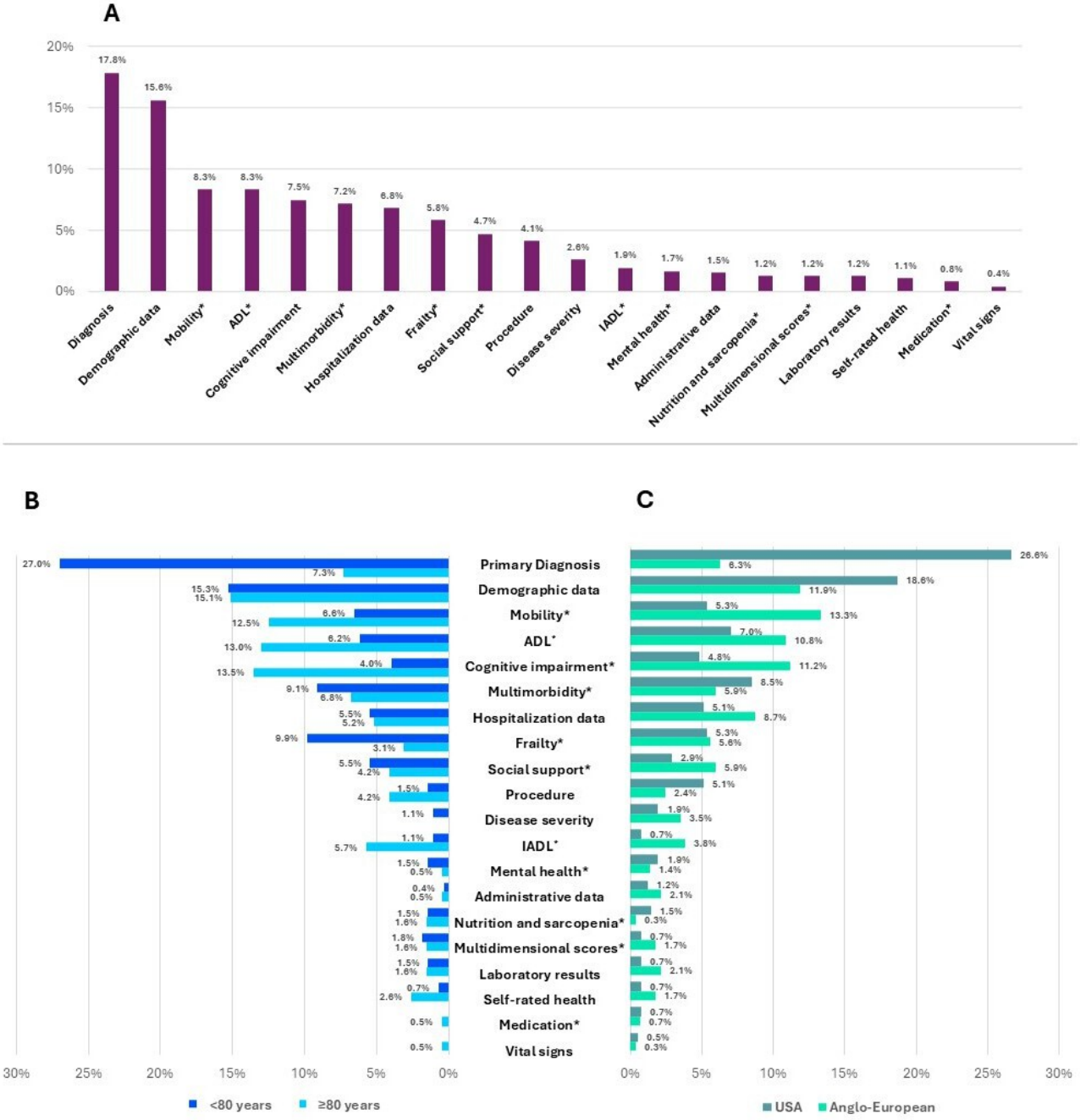
## Predictor Domains

After initial data inspection, 20 predictor domains representing common health care data and geriatric syndromes were defined, and all the extracted predictors were categorized accordingly. Diagnoses (129/723, 17.8%) and demographic data (113/723, 15.6%) formed the largest proportion. The assessment domains of geriatric syndromes, such as mobility (60/723, 8%), ADL (60/723, 8%), cognitive impairment (54/723, 8%), and frailty (42/723, 6%), were commonly represented as well (Figure 3A). Geriatric syndromes were more often predictive in the oldest old category (all: 352/723, 48.7%; <80 y: 129/274, 47.1%; ≥80 y: 121/192, 63%). In contrast, the frailty domain was a more frequent predictor in the younger old category (<80 y: 27/274, 10% vs ≥80 y: 6/192, 3%), as shown in Figure 3B.



XSL•FO
RenderX

**Figure 3.** Predictor domains. (A) Dark red columns: proportion of predictors across predictor domains (n=723). (B) Proportion of predictors across all predictor domains stratified by age. Dark blue bars: mean age <80 years (n=274); light blue bars: mean age ≥80 years (n=192). (C) Proportion of predictors across all predictor domains stratified by study origin. Dark green bars: US origin (n=413); light green bars: Anglo-European origin (n=286). ADL: activities of daily living, IADL: instrumental activities of daily living. *Geriatric syndromes.




Geriatric syndromes were more often extracted from Anglo-European compared to US studies (USA: 163/413, 39.5%; Anglo-European: 174/286, 60.8%). Among these, mobility (38/286, 13%), demographic data (34/286, 12%), along with cognitive impairment (32/286, 11%), and ADL (31/286, 11%) emerged as the predominant predictor domains (Figure 3C).

In total, 93 combinations of domain sets could be identified in the literature in our review. Contrasting these sets with all potential predictor combinations of the 20 domains, the greatest concordance with the highest average Jaccard Index score of 0.28 was found for the 8 elements shown in the upper part of Table 3 as a potentially generalizable domain set. As many literature-derived sets comprise solely 1 element, based on an 8-element set, the best Jaccard Index of 2 one-element sets could be 0.125. Thus, achieving a higher index value of 0.28 with an 8-element set across all literature represents a favorable outcome.



XSL•FO
RenderX

**Table .** Predictor domains most concordant with the literature and potentially generalizable. Upper part: Domain set with the highest mean Jaccard index in comparison to all sets found in the literature, in alphabetical order, including the top 3 predictors of each domain. Lower part: Other domains recognized in the review, in alphabetical order, including the top 3 predictors of each domain.

| Domain | Predictor examples |
|---|---|
| Domain set most concordant with the literature | |
| ADL[a] | ADL NSS[b], Barthel index [27], Katz score [25] |
| Cognitive impairment | Dementia diagnosis, confusion assessment method [33], short portable mental status questionnaire [34] |
| Demographic data | Age, sex (female), ethnicity |
| Hospitalization data | Length of stay, ICU[c] treatment, source of admission |
| Primary diagnosis | Wound problems, cardiopulmonary disease, infectious disease |
| Mobility | Falls, De Morton Mobility Index [30], pre-fracture mobility score [35] |
| Multimorbidity | American Society of Anesthesiologists score [24], number of comorbidities, Charlson comorbidity index [26] |
| Social support | Living alone or with companion, preadmission professional support, caregiver support |
| Other domains | |
| Administrative data | Level of care hospital, prior hospitalizations, prior ICU admission |
| Disease severity | Glasgow coma scale [36], injury severity score [28], injury severity NSS |
| IADL[d] | IADL no specific score, Lawton index [37], InterRAI Acute Care (finances) [38] |
| Frailty | Frailty index[e], clinical frailty scale [39], Fried frailty scale [40] |
| Laboratory results | Anemia, fluid and electrolyte disorder, hypoalbuminemia |
| Medication | Number of medication, vitamin K antagonist therapy, therapeutic anticoagulation |
| Mental health | Geriatric depression scale [41], depression diagnosis, 8-item Patient Health Questionnaire [42] |
| Multidimensional scores | Hospital Admission Risk Profile [43], ISAR[f] (≥2) [44] and CGA[g] normal, multidimensional prognostic index (every 0.1 increase) [45] |
| Nutrition and sarcopenia | BMI [46], Mini Nutrition Assessment short form [47], sarcopenia diagnosis |
| Procedures | Orthopedic or trauma surgery, anesthesia or ICU treatment procedure, gastrointestinal surgery |
| Self-rated health | Older people's quality of life questionnaire [32], 3-item brief health literacy screen, Short Form-12 physical component summary [48] |
| Vital signs | Vital capacity, respiratory rate, systolic blood pressure |

[a]ADL: activities of daily living.

[b]NSS: no specific score.

[c]ICU: intensive care unit.

[d]IADL: instrumental activities of daily living.

[e]Frailty index: different assessments possible.

[f]ISAR: identification of seniors at risk.

[g]CGA: comprehensive geriatric assessment.

## Qualitative Studies

A total of 10 studies were found that mostly utilized semistructured interviews complemented by observational techniques and reviews of patients' clinical records. Participants encompassed patients (n=8, 80%), health care professionals (n=8, 80%), and informal caregivers or relatives (n=2, 20%).

All study characteristics can be found in Multimedia Appendix 2.

A total of 98 predictors supporting an appropriate discharge decision were extracted. Predictors spanned various discharge destinations, namely discharge in general (n=36, 37%), return to home (n=25, 26%), transfer to rehabilitation (n=15, 15%), SNF (n=14, 14%), nursing home (n=5, 5%), and home with support (n=3, 3%).



XSL•FO
RenderX

These 98 predictors were categorized into 10 domains aligning partially with those identified in the quantitative studies: patient or caregiver involvement (n=27, 28%), organizational structures (n=16, 16%), health status or morbidity (n=16, 16%), communication among health care professionals or providers (n=9, 9%), social support (n=7, 7%), staff education (n=7, 7%), regional aspects (n=6, 6%), administrative health care data (n=4, 4%), hospitalization data (n=4, 4%), and home assessment (n=2, 2%).

Within the domain of patient or caregiver involvement, effective strategies emphasize communication and respect. Notable recurring themes include discussing discharge plans with patients and caregivers in a timely manner, addressing their capability to self-manage post-discharge, providing detailed medication schedules, and honoring patient preferences. As highlighted in 1 study, "Patients reported receiving contradictory information, especially with respect to medications and how to manage their care following discharge" [49], whereas another underscored that "...elders who thought they received enough information about how to manage their care after they left the hospital reported feeling satisfied" [50].

Regarding organizational structures, clear responsibilities, including staff continuity, geriatric comanagement, and standardized discharge procedures, emerged as significant components facilitating discharge decisions. A recent study emphasized the "...lack of recognized decision-making tools or algorithms as a critical issue in practice" [51].

The domain health status or morbidity includes parameters akin to those observed in quantitative studies, including ADL and cognitive impairment due to dementia or delirium. In 1 study, it was cited that "... clinicians noted the clinical challenges of managing delirium and the need for development of patient pathways for those with delirium" [52].

# Discussion

## Principal Findings

This scoping review identified and synthesized predictors of appropriate PAC destinations for older adults (≥65 y), transitioning from acute care to specific discharge destinations. Based on the analysis of 181 studies (171 quantitative, 10 qualitative), key predictors varied by discharge destination. Diagnosed dementia and deficits in instrumental activities of daily living were frequently associated with discharge to nursing homes, while longer length of stay and cardiopulmonary disease predicted discharge to rehabilitation, and good performance in activities of daily living and mobility assessments favored discharge home without support. Furthermore, the review highlighted the influence of age and geographical origin, with geriatric syndromes being more prominent in those aged 80 years and older and in non-US studies. The 8 predictor domains—cognitive impairment, activities of daily living, demographics, social support, hospitalization data, multimorbidity, mobility, and primary diagnosis—demonstrated the highest concordance with existing literature. These findings provide a comprehensive overview to inform clinical decision-making and the development of clinical decision support systems, including machine learning apps, aimed at optimizing PAC planning for older adults.

## Comparison to Prior Work

Age emerged as the most frequent predictor over nearly all subgroups. Older adults ≥80 years tend to be more often discharged to nursing homes rather than rehabilitation facilities [53]. This could be due to the reduced availability and accessibility of facilities as well as misconceptions regarding their potential for improvement through these services. Investigations involving populations with a mean age of ≥80 years demonstrate an increased use of geriatric syndromes as predictors for discharge destinations, identifiable by a comprehensive geriatric assessment (CGA) [54]. Notably, frailty appears more frequently as a predictive factor in studies encompassing individuals with a mean age of <80 years. This could be because frailty was not assessed in these studies or because frailty has been mediated by age or functional parameters such as the BI [55,56]. Qualitative analysis revealed that patient or caregiver involvement was another important predictor for appropriate discharge destination, although hardly used in quantitative studies.

To address the observed heterogeneity in predictors, our review presents a higher-level categorization based on predictor domains. This provides a more standardized and manageable framework for analysis. A CGA is another established method for capturing the heterogeneity of potential limitations across geriatric domains, thereby improving the likelihood of living at home after 3 months [57]. Comparing the predictor domains most concordant with existing literature with a CGA, it stands out that 5 of the 8 predictor domains (mobility, ADL, multimorbidity, social support, and cognitive impairment) are also core elements of a CGA [54]. Therefore, a CGA supplemented by routine measurements such as current hospitalization, primary diagnosis, and demographic data forms a solid foundation for enhancing prediction accuracy of the most appropriate discharge setting and is probably one reason why it ultimately reduces discharge to higher levels of care [58].

Discharge destination choices are inherently contingent upon local health care infrastructure and accessible resources. Therefore, our analysis distinguishes between Anglo-European studies, predominantly characterized by publicly funded health care systems, and US studies with a higher proportion of pay-per-service options. Anglo-European research tends to prioritize geriatric syndromes and length of hospital stay as key predictors, whereas US studies place greater emphasis on diagnostic factors. Variations in health care systems result in disparate PAC options whose definitions diverge across settings [59,60]. For example, we found many predictors for discharge to SNF, uniquely offered in the United States, combining skilled nursing care with rehabilitative interventions, even though similar initiatives are emerging in European countries such as Germany [3,61]. Despite regional differences, parallel options exist for clinically stable yet functionally declined older patients requiring both nursing care and rehabilitation within other health care frameworks, illustrated by examples like the Australian respite residential aged care or early inpatient geriatric rehabilitation in Germany [6,62]. Regrettably, we did not find



XSL•FO
RenderX

any comparative analyses examining various PAC alternatives across countries.

A lack of standardized decision support tools assisting health care professionals when deciding on the best discharge option is one of the issues that were raised in the qualitative studies [51,63,64]. This was also supported by a recent scoping review by Singh et al [65] focusing on digital health solutions facilitating transitions in care. In a systematic review on 35 CDSSs for PAC referral that included mainly studies of non-geriatric adults, Kennedy et al [13] revealed that merely 14% have been integrated into regular clinical practice, potentially hindered by constraints in time resources. Conversely, positive outcomes have emerged from specific implementations, such as a pre- or post-implementation study of a CDSS with a 2-step approach that showed a significant reduction of readmission rates [66]. Also, other binomial prediction models for discharge destinations, for example, for routine versus nonroutine discharge as in Karhade et al [67], showed promising results with an area under the curve of 0.823 in a machine learning model. However, when comparing these findings to one of the few CDSS developed for multinomial differentiation across 6 discharge destinations involving over 14,000 participants, it achieved a lower overall area under the curve of 0.685 [68]. The findings presented in this review will help to develop future multinomial COC or discharge prediction models by identifying eligible features.

### Strengths and Limitations

A key strength of our scoping review lies in the detailed differentiation of discrete prediction parameters, which enabled us to identify and statistically analyze a robust set of core predictor domains. However, our review is not without limitations, which can be categorized as pertaining to the included studies themselves and our review methodology.

We acknowledge the following limitations. The first limitation was the predominance of studies conducted in the US health care system with its specific PAC options and at least partially unmet health care needs due to some patients' lack of insurance coverage [69,70]. To address this limitation, we compared US studies with studies from Europe, Canada, and Australia with more similarly conceptualized health care systems, even though we are aware that this simplifies the complex differences between countries in the Anglo-European group that also exist.

The second limitation was the frequent use of nonstandardized assessments that we addressed by summarizing these measures into a "no specific score" predictor. For example, the predictors "problems with personal hygiene" or "ADL no specific score" are included in both the Katz Score and BI that again report comparable information. This indicates a need for a wider use of validated assessment instruments to capture the functional status in older patients in a uniform way. When selecting assessments for the future development of a CDSS tool or compiling assessments for a CGA, these validated and established assessments should be used.

Third, many of the included studies presented composite end points of different discharge destinations with unknown proportions among the different destinations as well as different populations concerning investigated diseases and clinical settings, which limits the validity of the extracted predictors for generalizable discharge destination options. Wherever possible, we tried to extract the singular discharge destination.

Fourth, focusing on the older geriatric population might have led to an omission of other studies mainly including nongeriatric adults that might also have included valuable information on the topic. For example, a review by Kennedy et al [13] found 33 studies presenting CDSS for optimizing discharge destinations, but only 6 studies were conducted in older age populations.

Fifth, our focus on positive predictors to avoid often confusing directions of effect among the negative predictors may have resulted in an undercapture of relevant predictors. With the negative predictors representing only 15.5% of the predictors, we consider the bias of this approach to be low.

Sixth, there is a lack of further details on the extracted predictors, for example, concerning the timepoint of the assessment or the direction of effect of the predictor. These parameters have been found relevant in other studies [66]. We partially addressed this lack of detail by manually extracting the direction of effect of the most frequent predictors stratified according to different discharge destinations. The aim of a scoping review is to cover a broader topic. This approach, however, results in a large heterogeneity among the included studies and a lack of details. Thus, the analysis provided here remains on a descriptive level and offers the opportunity for future research.

Seventh, only 10 studies were included in the qualitative analysis, with most studies from the 1990s and early 2000s, limiting the informative value of the qualitative evidence synthesis.

Eighth, the broad scope of studies screened and the comprehensive analysis undertaken resulted in a literature search limited to publications up to 2022. Nevertheless, we are confident that our review of this relevant, yet relatively stable, topic remains current.

### Conclusion and Implications

This scoping review synthesized evidence on predictors of appropriate discharge destinations for older adults, highlighting the heterogeneity of factors influencing care transitions. The identified 8-domain set—including demographic data, activities of daily living, social support, hospitalization data, primary diagnosis, cognitive impairment, mobility, and multimorbidity—underscores the value of comprehensive geriatric assessment in guiding discharge planning. This work will directly inform the feature selection process for a machine learning algorithm within the SURGE-Ahead project, designed to improve discharge recommendations. Recognizing the limitations of existing data and the need for system-specific adaptation, we advocate for continued research and the implementation of evidence-based discharge planning strategies.



XSL•FO
RenderX

## Acknowledgments

We thank Prof Dr Karen Andersen-Ranberg for critical discussion of preliminary data.

The SURGE-Ahead Study Group comprises Karen Andersen-Ranberg, Sophia Andres, Sibylle Beck, Christian Bolenz, Simone Brefka, Raffael Cintean, Karen Clauss, Dhayana Dallmeier, Genia Decker, Miriam Deniz, Michael Denkinger, Vanessa Disque, Anna Lena Flagmeier, Marina Liselotte Fotteler, Florian Gebhard, Julia Geiger, Janina Hahn, David Hansen, Thomas Hoffmann, Wolfgang Janni, Hans Kestler, Reinhold Kilian, Jochen Klenk, Thomas Derya Kocar, Aleyna Koese, Johann Kraus, Christoph Leinert, Elena Leinert, Fabia Mangold, Judith Mantz, Annabel Müller-Stierlin, Gabi Müller, Nadir Nasir, Graziano Onder, Desmond O'Neill, Marcin Orzechowski, Carlos Pankratz, Nina Parchmann, Nuh Rahbari, Johannes Schobel, Fabienne Schochter, Konrad Schütze, Tobias Skuban-Eiseler, Vytautas Stasiunaitis, Florian Steger, Walter Swoboda, Adriane Uihlein, Sebastian Voigt-Radloff, Martin Wehling, Pascal Wendel, Felix Wezel, Jana Willems, Jessica Wolf, Philip Wolf, Friedemann Zengerling, and Marco Zeptner.

Generative artificial intelligence was used to improve the quality of writing. It assisted with proofreading and enhancing proficiency in professional English. The model used is available at [71]. The models were run locally on an RTX A6000 GPU.

## Funding

This work was funded by the Federal Ministry of Education and Research (BMBF) and is part of the SURGE-Ahead—Supporting Surgery with Geriatric Co-Management and AI project (funding number: 01GY2101). The funder had no role in the study design, data extraction and synthesis, decision to publish, or preparation of the manuscript. CL, MLF, and TDK are employed in the SURGE-Ahead project.

## Authors' Contributions

Conceptualization: CL (lead), DD, KG, MD, MLF (equal), TDK (supporting)
Data curation: CL (lead), JW (supporting), KG, LB, MLF (supporting)
Formal analysis: CL (lead), MLF, TDK (supporting)
Funding acquisition: MD
Investigation: CL (lead), JW (supporting), KG, LB, MLF (equal)
Methodology: CL (lead), DD, KG, MD, MLF (equal), TDK (supporting)
Project administration: CL, MD (supporting), MLF (lead), TDK
Resources: CL (lead), MLF (supporting)
Supervision: DD (supporting), MD (lead)
Validation: MD (lead), TDK (supporting)
Visualization: CL (lead), MLF, TDK (supporting)
Writing – original draft: CL (lead), MLF (supporting)
Writing – review & editing: DD, JW, KG, LB (supporting), MD (lead), TDK (equal)
Final approval of the manuscript: all authors.
All the authors fulfill the International Committee of Medical Journal Editors criteria for authorship.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strings.
[PDF File, 217 KB - i-jmr_v15i1e76582_app1.pdf ]

Multimedia Appendix 2
Overview of included studies and reference list. (A) Overview of included quantitative studies ordered by year of publication. P: prospective observational study (eg, cohort study); R: retrospective observational study (eg, population-based study); ES: expert survey. (B) Overview of included qualitative studies ordered by year of publication.
[PDF File, 628 KB - i-jmr_v15i1e76582_app2.pdf ]

Multimedia Appendix 3
Predictors positive and negative stratified according to age, origin, study design, and discharge destinations.
[XLSX File, 663 KB - i-jmr_v15i1e76582_app3.xlsx ]

Multimedia Appendix 4



XSL•FO
RenderX

Predictors positive and negative, domain, discharge destinations, predictive strength indicators, and corresponding literature references.
[XLSX File, 88 KB - i-jmr_v15i1e76582_app4.xlsx ]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 122 KB - i-jmr_v15i1e76582_app5.pdf ]

## References


1. Wang YC, Chou MY, Liang CK, Peng LN, Chen LK, Loh CH. Post-acute care as a key component in a healthcare system for older adults. Ann Geriatr Med Res 2019 Jun;23(2):54-62. [doi: 10.4235/agmr.19.0009] [Medline: 32743289]
2. Gonçalves-Bradley DC, Lannin NA, Clemson L, Cameron ID, Shepperd S. Discharge planning from hospital. Cochrane Database Syst Rev 2022 Feb 24;2(2):CD000313. [doi: 10.1002/14651858.CD000313.pub6] [Medline: 35199849]
3. Heiks C, Sabine N. Long term care and skilled nursing facilities. Dela J Public Health 2022 Dec;8(5):144-149. [doi: 10.32481/djph.2022.12.032] [Medline: 36751604]
4. Bachmann S, Finger C, Huss A, Egger M, Stuck AE, Clough-Gorr KM. Inpatient rehabilitation specifically designed for geriatric patients: systematic review and meta-analysis of randomised controlled trials. BMJ 2010 Apr 20;340(apr20 2):c1718. [doi: 10.1136/bmj.c1718] [Medline: 20406866]
5. Zimmerman S, Sloane PD. Definition and classification of assisted living. Gerontologist 2007;47(suppl 1):33-39. [doi: 10.1093/geront/47.supplement_1.33] [Medline: 18162567]
6. Kwetkat A, Lehmann T, Wittrich A. Early geriatric rehabilitation: an opportunity for the oldest old. Z Gerontol Geriatr 2014 Jul;47(5):372-378. [doi: 10.1007/s00391-014-0660-7] [Medline: 24906436]
7. Continuity and coordination of care: a practice brief to support implementation of the WHO framework on integrated people-centred health services. World Health Organization. 2018. URL: https://iris.who.int/handle/10665/274628 [accessed 2025-12-23]
8. Bahr SJ, Weiss ME. Clarifying model for continuity of care: a concept analysis. Int J Nurs Pract 2019 Apr;25(2):e12704. [doi: 10.1111/ijn.12704] [Medline: 30393894]
9. Uijen AA, Schers HJ, Schellevis FG, van den Bosch W. How unique is continuity of care? A review of continuity and related concepts. Fam Pract 2012 Jun;29(3):264-271. [doi: 10.1093/fampra/cmr104] [Medline: 22045931]
10. Hirschman KB, Shaid E, McCauley K, Pauly MV, Naylor MD. Continuity of care: the transitional care model. Online J Issues Nurs 2015 Sep 30;20(3):1. [Medline: 26882510]
11. Kind AJH, Smith MA, Frytak JR, Finch MD. Bouncing back: patterns and predictors of complicated transitions 30 days after hospitalization for acute ischemic stroke. J American Geriatrics Society 2007 Mar;55(3):365-373. [doi: 10.1111/j.1532-5415.2007.01091.x]
12. Eichler S, Völler H, Reibis R, et al. Geriatric or cardiac rehabilitation? Predictors of treatment pathways in advanced age patients after transcatheter aortic valve implantation. BMC Cardiovasc Disord 2020 Apr 6;20(1):158. [doi: 10.1186/s12872-020-01452-x] [Medline: 32252646]
13. Kennedy EE, Bowles KH, Aryal S. Systematic review of prediction models for postacute care destination decision-making. J Am Med Inform Assoc 2022 Dec 28;29(1):176-186. [doi: 10.1093/jamia/ocab197] [Medline: 34757383]
14. Leinert C, Fotteler M, Kocar TD, et al. Supporting SURgery with GEriatric Co-management and AI (SURGE-Ahead): a study protocol for the development of a digital geriatrician. PLoS One 2023;18(6):e0287230. [doi: 10.1371/journal.pone.0287230] [Medline: 37327245]
15. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
16. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
17. Leinert C, Fotteler M, Kocar TD, Dallmeier D, Denkinger M. Predictors for continuity of care decisions for older inpatients in acute care settings: a scoping review. OSF. 2022 Mar 22. URL: https://osf.io/6qsvb/ [accessed 2025-12-15]
18. Hirshon JM, Risko N, Calvello EJB, et al. Health systems and services: the role of acute care. Bull World Health Organ 2013 May 1;91(5):386-388. [doi: 10.2471/BLT.12.112664] [Medline: 23678202]
19. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8(1):45. [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
20. Palm W, Webb E, Hernández-Quevedo C, et al. Gaps in coverage and access in the European Union. Health Policy 2021 Mar;125(3):341-350. [doi: 10.1016/j.healthpol.2020.12.011] [Medline: 33431257]
21. Ridic G, Gleason S, Ridic O. Comparisons of health care systems in the United States, Germany and Canada. Mater Socio Med 2012;24(2):112-120. [doi: 10.5455/msm.2012.24.112-120] [Medline: 23678317]
22. Dixit SK, Sambasivan M. A review of the Australian healthcare system: a policy perspective. SAGE Open Med 2018;6:2050312118769211. [doi: 10.1177/2050312118769211] [Medline: 29686869]




XSL•FO
RenderX

23. Ahmadinejad N, Chung Y, Liu L. J-score: a robust measure of clustering accuracy. PeerJ Comput Sci 2023;9:e1545. [doi: 10.7717/peerj-cs.1545] [Medline: 37705621]
24. Knuf KM, Maani CV, Cummings AK. Clinical agreement in the American Society of Anesthesiologists physical status classification. Perioper Med (Lond) 2018;7:14. [doi: 10.1186/s13741-018-0094-7] [Medline: 29946447]
25. Katz S, Downs TD, Cash HR, Grotz RC. Progress in development of the index of ADL. Gerontologist 1970;10(1):20-30. [doi: 10.1093/geront/10.1_part_1.20] [Medline: 5420677]
26. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
27. Mahoney FI, Barthel DW. Functional evaluation: the Barthel index: a simple index of independence useful in scoring improvement in the rehabilitation of the chronically ill. Md State Med J 1965;14:61-65. [Medline: 14258950]
28. Baker SP, O'Neill B, Haddon W Jr, Long WB. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. J Trauma 1974 Mar;14(3):187-196. [Medline: 4814394]
29. Foss NB, Kristensen MT, Kehlet H. Prediction of postoperative morbidity, mortality and rehabilitation in hip fracture patients: the cumulated ambulation score. Clin Rehabil 2006 Aug;20(8):701-708. [doi: 10.1191/0269215506cre987oa] [Medline: 16944827]
30. de Morton NA, Davidson M, Keating JL. The de Morton Mobility Index (DEMMI): an essential health index for an ageing world. Health Qual Life Outcomes 2008 Aug 19;6(1):63. [doi: 10.1186/1477-7525-6-63] [Medline: 18713451]
31. Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol 1994 Mar;49(2):M85-M94. [doi: 10.1093/geronj/49.2.m85] [Medline: 8126356]
32. Bowling A. The psychometric properties of the older people's quality of life questionnaire, compared with the CASP-19 and the WHOQOL-OLD. Curr Gerontol Geriatr Res 2009;2009:298950. [doi: 10.1155/2009/298950] [Medline: 20168974]
33. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. Ann Intern Med 1990 Dec 15;113(12):941-948. [doi: 10.7326/0003-4819-113-12-941] [Medline: 2240918]
34. Erkinjuntti T, Sulkava R, Wikström J, Autio L. Short Portable Mental Status Questionnaire as a screening test for dementia and delirium among the elderly. J Am Geriatr Soc 1987 May;35(5):412-416. [doi: 10.1111/j.1532-5415.1987.tb04662.x] [Medline: 3571790]
35. Voeten SC, Nijmeijer WS, Vermeer M, Schipper IB, Hegeman JH, DHFA Taskforce study group. Validation of the fracture mobility score against the parker mobility score in hip fracture patients. Injury 2020 Feb;51(2):395-399. [doi: 10.1016/j.injury.2019.10.035] [Medline: 31668574]
36. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. Lancet 1974 Jul 13;2(7872):81-84. [doi: 10.1016/s0140-6736(74)91639-0] [Medline: 4136544]
37. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969;9(3):179-186. [doi: 10.1093/geront/9.3_Part_1.179] [Medline: 5349366]
38. Gray LC, Beattie E, Boscart VM, et al. Development and testing of the interRAI acute care: a standardized assessment administered by nurses for patients admitted to acute care. Health Serv Insights 2018;11:1178632918818836. [doi: 10.1177/1178632918818836] [Medline: 30618486]
39. Cheung A, Haas B, Ringer TJ, McFarlan A, Wong CL. Canadian study of health and aging clinical frailty scale: does it predict adverse outcomes among geriatric trauma patients? J Am Coll Surg 2017 Nov;225(5):658-665. [doi: 10.1016/j.jamcollsurg.2017.08.008] [Medline: 28888692]
40. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001 Mar;56(3):M146-M156. [doi: 10.1093/gerona/56.3.m146] [Medline: 11253156]
41. Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1982;17(1):37-49. [doi: 10.1016/0022-3956(82)90033-4] [Medline: 7183759]
42. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
43. Sager MA, Rudberg MA, Jalaluddin M, et al. Hospital Admission Risk Profile (HARP): identifying older patients at risk for functional decline following acute medical illness and hospitalization. J Am Geriatr Soc 1996 Mar;44(3):251-257. [doi: 10.1111/j.1532-5415.1996.tb00910.x]
44. McCusker J, Bellavance F, Cardin S, Trépanier S, Verdon J, Ardman O. Detection of older people at increased risk of adverse health outcomes after an emergency visit: the ISAR screening tool. J Am Geriatr Soc 1999 Oct;47(10):1229-1237. [doi: 10.1111/j.1532-5415.1999.tb05204.x] [Medline: 10522957]
45. Pilotto A, Ferrucci L, Franceschi M, et al. Development and validation of a multidimensional prognostic index for one-year mortality from comprehensive geriatric assessment in hospitalized older patients. Rejuvenation Res 2008 Feb;11(1):151-161. [doi: 10.1089/rej.2007.0569] [Medline: 18173367]
46. Eknoyan G. Adolphe Quetelet (1796-1874)—the average man and indices of obesity. Nephrol Dial Transplant 2008 Jan;23(1):47-51. [doi: 10.1093/ndt/gfm517] [Medline: 17890752]




XSL•FO
RenderX

47. Kaiser MJ, Bauer JM, Ramsch C, et al. Validation of the Mini Nutritional Assessment short-form (MNA-SF): a practical tool for identification of nutritional status. J Nutr Health Aging 2009 Nov;13(9):782-788. [doi: 10.1007/s12603-009-0214-7] [Medline: 19812868]
48. Ware J, Kosinski M, Keller SD. A 12-Item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
49. Bull MJ, Kane RL. Gaps in discharge planning. J Appl Gerontol 1996 Dec;15(4):486-500. [doi: 10.1177/073346489601500406]
50. Bull MJ. Patients' and professionals' perceptions of quality in discharge planning. J Nurs Care Qual 1994 Jan;8(2):47-61. [doi: 10.1097/00001786-199401000-00009] [Medline: 8312594]
51. Bradley G, Baker K, Bailey C. Exploring how occupational therapists and physiotherapists evaluate rehabilitation potential of older people in acute care. Br J Occup Ther 2022 Mar;85(3):199-207. [doi: 10.1177/03080226211011386] [Medline: 40337082]
52. Isbel ST, Jamieson MI. Views from health professionals on accessing rehabilitation for people with dementia following a hip fracture. Dementia (London) 2017 Nov;16(8):1020-1031. [doi: 10.1177/1471301216631141] [Medline: 26843421]
53. Marengoni A, Agüero-Torres H, Timpini A, Cossi S, Fratiglioni L. Rehabilitation and nursing home admission after hospitalization in acute geriatric patients. J Am Med Dir Assoc 2008 May;9(4):265-270. [doi: 10.1016/j.jamda.2008.01.005] [Medline: 18457803]
54. Deutsche Gesellschaft für Geriatrie eV. (DGG). S3 guideline: comprehensive geriatric assessment (CGA) in hospitalized patients, long version 2.3. AWMF Leitlinien-Register. 2024. URL: https://register.awmf.org/de/leitlinien/detail/084-003 [accessed 2024-11-26]
55. Hegendörfer E, Vaes B, Van Pottelbergh G, Matheï C, Verbakel J, Degryse JM. Predictive accuracy of frailty tools for adverse outcomes in a cohort of adults 80 years and older: a decision curve analysis. J Am Med Dir Assoc 2020 Mar;21(3):440. [doi: 10.1016/j.jamda.2019.08.029] [Medline: 31678074]
56. Gordon EH, Peel NM, Hubbard RE, Reid N. Frailty in younger adults in hospital. QJM-Int J Med 2023 Oct 23;116(10):845-849. [doi: 10.1093/qjmed/hcad173] [Medline: 37467071]
57. Ellis G, Gardner M, Tsiachristas A, et al. Comprehensive geriatric assessment for older adults admitted to hospital. Cochrane Database Syst Rev 2017 Sep 12;9(9):CD006211. [doi: 10.1002/14651858.CD006211.pub3] [Medline: 28898390]
58. Eamer G, Taheri A, Chen SS, et al. Comprehensive geriatric assessment for older people admitted to a surgical service. Cochrane Database Syst Rev 2018 Jan 31;1(1):CD012485. [doi: 10.1002/14651858.CD012485.pub2] [Medline: 29385235]
59. van Balen R, Gordon AL, Schols J, Drewes YM, Achterberg WP. What is geriatric rehabilitation and how should it be organized? A Delphi study aimed at reaching European consensus. Eur Geriatr Med 2019 Dec;10(6):977-987. [doi: 10.1007/s41999-019-00244-7] [Medline: 34652774]
60. Sezgin D, O'Caoimh R, O'Donovan MR, et al. Defining the characteristics of intermediate care models including transitional care: an international Delphi study. Aging Clin Exp Res 2020 Nov;32(11):2399-2410. [doi: 10.1007/s40520-020-01579-z] [Medline: 32430887]
61. Klauber J, Wasem J, Beivers A, Mostert C, Scheller-Kreinsen D, editors. Krankenhaus-Report 2024: Strukturreform: Springer; 2024. [doi: 10.1007/978-3-662-68792-5]
62. Mitchell R, Harvey L, Draper B, Brodaty H, Close J. Risk factors associated with residential aged care, respite and transitional aged care admission for older people following an injury-related hospitalisation. Arch Gerontol Geriatr 2017 Sep;72:59-66. [doi: 10.1016/j.archger.2017.05.012] [Medline: 28599139]
63. Burke RE, Lawrence E, Ladebue A, et al. How hospital clinicians select patients for skilled nursing facilities. J Am Geriatr Soc 2017 Nov;65(11):2466-2472. [doi: 10.1111/jgs.14954] [Medline: 28682456]
64. McWilliam CL, Sangster JF. Managing patient discharge to home: the challenges of achieving quality of care. Int J Qual Health Care 1994 Jun;6(2):147-161. [doi: 10.1093/intqhc/6.2.147] [Medline: 7953214]
65. Singh H, Tang T, Gray CS, et al. Recommendations for the design and delivery of transitions-focused digital health interventions: rapid review. JMIR Aging 2022 May 19;5(2):e35929. [doi: 10.2196/35929] [Medline: 35587874]
66. Bowles KH, Ratcliffe SJ, Holmes JH, et al. Using a decision support algorithm for referrals to post-acute care. J Am Med Dir Assoc 2019 Apr;20(4):408-413. [doi: 10.1016/j.jamda.2018.08.016] [Medline: 30414821]
67. Karhade AV, Ogink P, Thio Q, et al. Development of machine learning algorithms for prediction of discharge disposition after elective inpatient surgery for lumbar degenerative disc disorders. Neurosurg Focus 2018 Nov 1;45(5):E6. [doi: 10.3171/2018.8.FOCUS18340] [Medline: 30453463]
68. Sultana I, Erraguntla M, Kum HC, Delen D, Lawley M. Post-acute care referral in United States of America: a multiregional study of factors associated with referral destination in a cohort of patients with coronary artery bypass graft or valve replacement. BMC Med Inform Decis Mak 2019 Nov 14;19(1):223. [doi: 10.1186/s12911-019-0955-0] [Medline: 31727058]
69. Rice T, Rosenau P, Unruh LY, Barnes AJ. United States: health system review. Health Syst Transit 2020 Dec;22(4):1-441. [Medline: 33527901]
70. Kennedy EE, Davoudi A, Hwang S, et al. Identifying barriers to post-acute care referral and characterizing negative patient preferences among hospitalized older adults using natural language processing. AMIA Annu Symp Proc 2023;2022:606-615. [Medline: 37128417]




XSL•FO
RenderX

71. Saama AI Labs. Llama3-OpenBioLLM-70B. Hugging Face. URL: https://huggingface.co/aaditya/Llama3-OpenBioLLM-70B [accessed 2025-12-15]

## Abbreviations

**ADL:** activities of daily living
**BI:** Barthel index
**CDSS:** clinical decision support systems
**CGA:** Comprehensive geriatric assessment
**COC:** continuity of care
**IADL:** instrumental activities of daily living
**PAC:** postacute care
**SNF:** skilled nursing facility
**SURGE-Ahead:** Supporting Surgery with Geriatric Co-Management and AI

*Edited by N Cahill; submitted 26.Apr.2025; peer-reviewed by K Bowles, M Gagesch; revised version received 30.Sep.2025; accepted 08.Oct.2025; published 22.Jan.2026.*

*Please cite as:*
*Leinert C, Fotteler ML, Kocar TD, Wolf J, Beissel L, Grummich K, Dallmeier D, Denkinger M*
*Identifying Key Predictors of Appropriate Discharge Destinations for Older Inpatients in Acute Care: Scoping Review*
*Interact J Med Res 2026;15:e76582*
*URL: https://www.i-jmr.org/2026/1/e76582*
*doi:10.2196/76582*

© Christoph Leinert, Marina Liselotte Fotteler, Thomas Derya Kocar, Jessica Wolf, Lisa Beissel, Kathrin Grummich, Dhayana Dallmeier, Michael Denkinger. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 22.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# The Effectiveness, Facilitators, and Barriers of Digital Mental Health Services for First Nations People in Australia: Systematic Scoping Review

Siyu Zhai[1]; Andrew Goodman[1,2]; Anthony C Smith[3,4,5,6]; Sandra Diminic[1,7]; Xiaoyun Zhou[1,7]

[1]School of Public Health, The University of Queensland, 266 Herston Road, Herston, Queensland, Australia

[2]Australian eHealth Research Centre (AEHRC), Commonwealth Scientific and Industrial Research Organisation (CSIRO), Herston, Australia

[3]Centre for Online Health, The University of Queensland, Brisbane, Australia

[4]Hans Christian Andersen Children's Hospital, Odense University Hospital, Odense, Denmark

[5]Centre for Innovative Medical Technology, University of Southern Denmark, Odense, Denmark

[6]Centre for Health Services Research, The University of Queensland, Herston, Queensland, Australia

[7]Queensland Centre for Mental Health Research, Brisbane, Australia

**Corresponding Author:**
Siyu Zhai
School of Public Health, The University of Queensland, 266 Herston Road, Herston, Queensland, Australia

## Abstract

**Background:** First Nations people in Australia experience inequitable mental health outcomes and service access. Digital mental health (DMH) services, which refer to offering mental health services through digital platforms, are considered potential solutions to address such mental health service inequities and improve the mental health outcomes of First Nations Australians. However, evidence on the effectiveness of DMH services for First Nations people in Australia is yet to be synthesized.

**Objective:** This systematic scoping review aims to fill this gap and to identify the facilitators and barriers that influence the implementation of DMH services among First Nations people in Australia.

**Methods:** A systematic search was conducted across 6 academic databases to search for studies related to DMH services for First Nations people in Australia. Search terms relating to First Nations people, geographic terminologies of Australia, mental health, and DMH services were used. Studies were included if they assessed the effectiveness of DMH services or the determinants of the facilitators and barriers of implementing DMH interventions among First Nations people in Australia. Data were extracted based on study design, targeted services, and research findings, and were then synthesized using a thematic analysis framework.

**Results:** In total, 22 studies met the inclusion criteria. DMH services were used to support and treat First Nations Australians and conduct psychological assessments in these individuals. Evidence of effectiveness was stronger for nonsevere mental health conditions. The determinants of the facilitators and barriers of the implementation of DMH services included the following: (1) organizational and administrative factors; (2) cultural appropriateness; (3) accessibility; (4) integration of DMH services in the existing health system; (5) engagement between clients and service providers; (6) coverage of different conditions and clients; (7) acceptability to DMH services; (8) digital literacy; and (9) efficiency.

**Conclusions:** Evidence on the use of DMH services for First Nations Australians remains heterogeneous in terms of study design and outcome measurement. DMH services appear to be most effective for managing nonsevere mental health conditions. Successful implementation requires multilevel structural support, including policy and organizational commitment, enhanced digital infrastructure, workforce training and engagement, and the design of culturally responsive DMH models to improve uptake and equitable access to mental health care among First Nations Australians.

**Trial Registration:** PROSPERO CRD42024612517; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024612517

***(Interact J Med Res 2026;15:e80386)*** doi:10.2196/80386

**KEYWORDS**

digital mental health; Aboriginal and Torres Strait Islander people; eHealth; mHealth; mobile health; telehealth; mental health services



XSL•FO
RenderX

## Introduction

Aboriginal and Torres Strait Islander people in Australia ("First Nations" people used hereafter) view mental health as a part of a holistic view of health across all life stages [1,2]. Maintaining a personal connection to traditional culture, health views, and community serves as an important protective factor for the mental health of First Nations people [1,3-6]. Therefore, reinforcing cultural strength and resilience would foster personal resilience and self-esteem, thereby protecting their individual mental health [7] and reducing the likelihood of developing mental illness [5,8,9]. Moreover, mental health services that acknowledge the right of self-determination and the need for cultural understanding of First Nations people [10], provide culturally appropriate care [11], or employ First Nations health workers [12] are also more likely to reinforce protective factors and achieve better treatment outcomes.

However, First Nations people are facing more complex mental health challenges than the general population in Australia due to external historical and social determinants [13]. Historically, colonization led to intergenerational trauma that worsened their mental health [14]. Currently, First Nations people continue to face daily discrimination [15], along with social inequity in areas such as education and employment [9]. These structural inequities increase the likelihood of mental health issues within these communities [16] and create mental health outcome gaps between First Nations people and the general population. For instance, although the national health survey showed that the prevalence of any mental illness among First Nations people dropped from 29.3% in 2014‐2015 [8] to 24% in 2018‐2019, it was still considered high [17]. Additionally, a survey conducted in regional Australia that applied structured clinical interviews in 2014‐2016 showed that the prevalence of any mental illness was 42.2% (4.2-fold higher than the prevalence in the general population) [5]. The prevalence of anxiety and mood disorders among First Nations people was estimated to be 1.6‐3.3 times the national prevalence in Australia [18]. Furthermore, 16% of deaths of First Nations people were related to suicide in 2001‐2005, which is 10% higher than the proportion in the general population [4].

Despite significant mental health needs and gaps, First Nations people experience barriers to accessing mental health services. First, structural racism, stigma toward mental health conditions, low health literacy, and less comfort in seeking help from health professionals prevent First Nations people from using more mental health services [11,12,19]. Second, a lack of local services and service providers creates barriers to travel for First Nations people [11,20], especially in areas with higher geographical remoteness that have a higher proportion of First Nations populations [21] but lower service access [22]. Moreover, community-controlled, culturally safe services may be unavailable for some First Nations people [23], and non-First Nations health workers who do not receive appropriate cultural safety training also contribute to service access barriers [20,24-26]. In 2018‐2019, only 31% of First Nations adults with high or very high levels of psychological distress received professional mental health services [27]. Moreover, mental health service gaps were mostly reported by First Nations primary health organizations (134 out of 198 organizations) [28]. Furthermore, among suicide cases, First Nations populations were approximately 2 times less likely than non-First Nations populations to have received mental health support (23.8% vs 43.3%) [29]. In the Northern Territory, mental health service utilization of First Nations youth is nearly half that of non-First Nations youth (1.9% vs 4.1%) [30].

Digital mental health (DMH) services refer to providing mental health services through digital technologies [31-33]. DMH services have been implemented for the prevention, screening, intervention, and rehabilitation of mental health conditions [34] with or without mental health professionals' guidance or support [35], and they demonstrate promising results for managing various mental health conditions, including anxiety disorders [36], depressive disorders [37], posttraumatic stress disorder [38], and alcohol and other drug-related mental disorders [39]. Moreover, DMH services have been evaluated among different populations, including children and young people [40], adults [41], and older populations [42].

DMH services have been evaluated among First Nations people worldwide [43-46] and have shown the potential to address the current health outcome gaps faced by them. Evidence from Canada, New Zealand, and the United States indicates that DMH services support improvements in psychological assessment [47-50] comparable to existing face-to-face services. DMH services were also found to improve clinical outcomes [49,51] and continuity in mental health care [52], both as stand-alone and blended care. International evidence also suggests that DMH services can reduce mental health gaps by improving the accessibility and availability of mental health services among First Nations people [23,34-36].

In Australia, evidence on DMH services among First Nations people is mixed. Although digital gaps remain a barrier to DMH implementation [53], government funding has enabled the provision of digital infrastructure in rural and remote First Nations communities [54]. Another service gap is the need for culturally appropriate services [23]. Evidence suggests that DMH services may fail to deliver culturally safe services [55,56] due to the preference of First Nations people for the familiarity and safety of face-to-face services [50]. However, some studies have indicated that digital health may support cultural safety by increasing the involvement of First Nations health workers [43,57] and community-led design services [58], and by allowing patients to see familiar faces via video conferencing [54]. DMH services also enable First Nations people with mental health needs to include family and community members in decision-making, without the need for travel [54].

Currently, most systematic reviews have focused on international evidence regarding the application of DMH services to First Nations people [43-46]. Existing Australian evidence is more likely to review broad and general health conditions rather than focus on mental health [47,54,57,59] or to solely focus on the design and assessment of DMH services for First Nations people rather than the implementation [58]. There is a paucity of evidence on the effectiveness of DMH services and the facilitators and barriers of implementing these services among First Nations people in Australia. Our study



XSL•FO
RenderX

aims to address this research gap by exploring the following research questions: (1) What is the effectiveness of DMH services for First Nations people in Australia? (2) What are the determinants of the facilitators and barriers of implementing DMH services among First Nations people in Australia? Such evidence is critical for policymakers, health providers, and First Nations users to make evidence-informed decisions while funding, implementing, prescribing, and choosing DMH services.

## Methods

### Reporting and Registration

The reporting of this study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines (Checklist 1) and was registered through PROSPERO (CRD42024612517).

### Acknowledgment to Indigenous Data Sovereignty

This study follows the Maiam Nayri Wingara principles [60]. The authors acknowledge that First Nations people have control of the data ecosystem; all data are available to First Nations people; and this study pays respect to First Nations communities, will be accountable to these communities, and empowers the sustainable self-determination of these communities.

### Outcome Definition

Our study outcomes include the effectiveness of DMH services and the determinants of facilitators and barriers. The effectiveness of DMH services is defined as the extent to which DMH services obtain the desired mental health outcomes. For example, for DMH services that are related to psychological diagnosis and assessment [49], the accuracy of diagnosis is considered as effectiveness. For treatments provided through DMH services [51,52,61], effectiveness is defined as the degree to which a treatment service improves mental health symptoms. The determinants of facilitators and barriers are defined as the determinants that can promote and impede, respectively, the implementation of DMH services for First Nations people in Australia.

### Eligibility Criteria

Based on the PICO (population, intervention, comparison, outcome) protocol, this review considered studies that (1) included participants who were First Nations Australians; (2) included interventions that were DMH services with various functions; (3) included comparisons that were typical mental health services or did not include comparisons; and (4) targeted outcomes that involved changes in mental health conditions. In addition, studies written in English were included. Studies that were irrelevant to our PICO criteria and those that did not have full text available were excluded. Moreover, there was no eligibility criterion relating to First Nations groups, date of publication, or age of participants. See Multimedia Appendix 1 for detailed inclusion and exclusion criteria and the PICO protocol.

### Search Strategy

A systematic search was conducted from September 9, 2024, to September 15, 2024, across 6 online databases: PsycINFO, PubMed, MEDLINE, Embase, Web of Science, and Google Scholar. Considering that more than 10,000 search results were obtained from Google Scholar, searching on this database was limited to the first 30 pages.

The search strategy was developed after consultation among 3 authors (SZ, XZ, and SD). Boolean operators, truncations, and search terms were tailored to each database. The search terms were organized around the following themes: First Nations people, geographic terminologies of Australia, mental health (including specific mental health conditions), and DMH services. Considering that DMH is a relatively new research area, there was no limitation on the publication date of the searched studies. See Multimedia Appendix 2 for details of the search strategy.

### Study Selection

Selected studies were imported into EndNote 20 and Covidence, and duplicates were removed at this stage. Afterwards, the titles and abstracts were screened. Studies found to be irrelevant to the review topic based on an assessment of the title and/or abstract were excluded. Studies that had insufficient information in the title and abstract were moved to full-text screening. During full-text screening, studies that met the exclusion criteria were excluded.

One reviewer (SZ) conducted the title and abstract screening. Uncertainty in this process was solved through discussion with another reviewer (XZ), and any disagreements were resolved by involving the third reviewer (SD). Two reviewers completed the full-text screening, where one reviewer (SZ) made initial decisions and the other reviewer (XZ) checked the findings. Therefore, the selection process was not blinded. Disagreements were resolved through discussion or by involving the third reviewer (SD).

### Data Extraction

Extracted data were tabulated based on author, year of publication, study design (qualitative, quantitative, or mixed methods), study aim, study population, research setting, type of DMH intervention, outcome (effectiveness, facilitators, and barriers), and key research findings. See Multimedia Appendix 3 for details of the extracted data.

For missing data, if the authors did not specify the reason, the reviewers tried to contact the authors to obtain relevant information. If such information was unavailable or no response was received, the missing data were excluded from the analysis process.

### Strategy of Data Synthesis

The synthesis of data followed the thematic analysis method introduced by Thomas and Harden [62] and was guided by the SWiM (Synthesis Without Meta-Analysis) guideline [63].

Reviewer SZ read selected studies completely and focused on the results section of those studies. Results relevant to the targeted outcomes of this review, such as the effect size of interventions or reported user experiences related to these



XSL•FO
RenderX

determinants, were summarized and coded into descriptive themes. Descriptive themes were categorized into subgroups and synthesized based on the type of data (qualitative or quantitative), research design (with or without a control group), and relevant determinants (facilitators, barriers, or effectiveness). Afterwards, reviewers standardized themes using the direction of the effect. Vote counting was conducted based on the direction of the effect. Then, the commonality among descriptive themes was found based on the direction of the effect or the different determinants they related to, which were used to formulate the analytical themes that summarized multiple similar descriptive themes. Analytical themes interpreted the primary results to fit the objectives of this review. To ensure the trustworthiness and validity of thematic analysis, during the generation of themes, reviewer SZ consulted reviewers XZ and SD to resolve any uncertainty. Once reviewer SZ completed generating themes, reviewer XZ checked all themes. Any disagreements were resolved by consulting the third analyst, SD.

### Quality Assessment

This study applied the Aboriginal and Torres Strait Islander Quality Appraisal Tool [64] to assess the quality of health research from a cultural sensitivity perspective. Reviewer SZ carefully reviewed the included studies and answered the 14 questions in the Aboriginal and Torres Strait Islander Quality Appraisal Tool based on the content of all selected studies. Moreover, as the instructions for this quality assessment tool suggested, the numbers of studies assessed as "yes," "partially," "no," and "unclear" have been reported. Reviewer SZ applied the Mixed Methods Appraisal Tool (MMAT) [65] for assessing the methodological quality of the included studies. Only the relevant part of the included study was assessed. For instance, if the qualitative part of a mixed-methods study was the only part relevant to this review, the study was appraised as a qualitative study. The MMAT does not recommend scoring each study; hence, this review only reports the performance of the included studies on each criterion and the overall trend of the design of the included studies. All quality assessment results were checked by reviewers XZ and SD. See Multimedia Appendices 4 and 5 and the Results section for the quality assessment results.

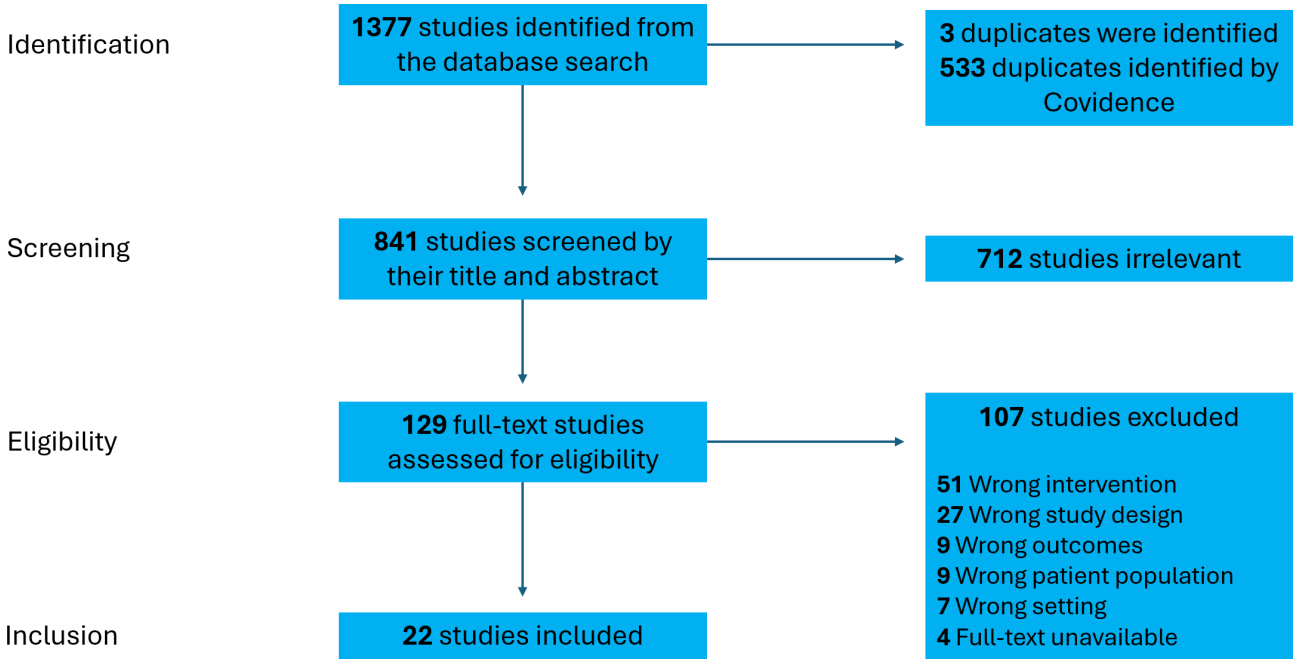
In this systematic scoping review, all studies were included regardless of quality rating. However, the quality appraisal results informed the interpretation of the searched studies. Considering this study was a narrative synthesis and did not include a meta-analysis, no statistical weighting or sensitivity analysis was performed.

## Results

### General Characteristics of the Studies

A total of 22 articles were included in this review [66-87]. The PRISMA flowchart is presented in Figure 1. See Multimedia Appendix 3 for detailed characteristics. Except for 1 study that solely reported barriers [73] and 4 studies that solely reported effectiveness [72,82,85,87], the rest of the studies reported more than one targeted outcome (ie, facilitators, barriers, and effectiveness). Moreover, 16 studies [67,68,71-74,76,77,79-84,86,87] were published within 5 years, and all included studies were published between 2015 and 2023.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of article screening and inclusion.




Among the included studies, there were 8 qualitative studies [66,67,69,74,75,78,79,83], 8 mixed-method studies [68,70,71,76,77,80,81,84], and 6 quantitative studies [72,73,82,85-87]. Among the quantitative studies, 3 adopted a cross-sectional research design [72,73,87], 1 was a crossover study [82], 1 was a randomized controlled trial [85], and 1 was a prospective cohort study [86].

All studies involved First Nations people as the clients of DMH services. There were 5 studies involving health care providers who provided DMH services to First Nations people



XSL•FO
RenderX

[66,69,78-80], 2 studies involving both clients and service providers [67,74], 12 studies involving adult clients or clients whose ages were unspecified [68,70-73,75,81,82,84-87], and 3 studies involving youth or student clients [76,77,83].

Regarding the types of DMH services, 5 studies were related to general perspectives on DMH services [66,77-80]. Among the studies, 2 involved DMH services delivered through web pages [70,86], 12 involved DMH services delivered through mobile apps [67-69,71,72,74-76,79,80,84,85], 5 involved DMH services delivered through telehealth methods (text or voice-only phone call) [70,73,86,87], 1 involved DMH services delivered through video conferencing [82], and 2 involved DMH services containing both online and offline activities [81,83]. See Table 1 and Multimedia Appendix 3 for the details of DMH service types.



XSL•FO
RenderX

**Table .** Overview of the findings in relation to the effectiveness of digital mental health services.

| Evidence type and study (author, year) | Data collection method | Data analysis method | Digital mental health service type | Major findings |
|---|---|---|---|---|
| Qualitative evidence | | | | |
| Dingwall et al [68], 2023 | Qualitative interviews at the end of the 4-week intervention | • Deductive approach for qualitative analysis | Aboriginal and Islander Mental Health Initiative for Youth (AIMhi-Y) app | • Participants had mental health improvements or gained help during tough times by using this app, which suggested its effectiveness. |
| Perdacher et al [74], 2022 | Semistructured interviews | • Thematic analysis under the framework provided by the constant comparison method | Stay Strong app (custodial version) | • Both clients and practitioners reported that the app improved the abilities of self-insight and self-reflection of clients and guided the positive thinking of clients.<br>• Both clients and practitioners reported that the app could enhance clients' confidence, view of self, and empowerment. |
| Povey et al [75], 2016 | Three focus groups | • Thematic analysis | AIMhi Stay Strong iPad app and the iBobbly suicide prevention app | • Both apps have the potential to help people with minor mental health conditions overcome difficulties.<br>• For the suicide prevention function, participants thought the app is useful for nonsevere conditions but may not be helpful in preventing suicide due to it being a severe condition. |
| Raphiphatthana et al [80], 2020 | Semistructured interviews | • Thematic analysis that oscillated between deductive and inductive approaches | General digital mental health services (but many of the responses were related to the use of the Stay Strong app) | • Electronic mental health was perceived as potentially beneficial and useful for First Nations communities. |
| Routledge et al [81], 2022 | Paper-and-pen test in classes or an online questionnaire | • Thematic analysis | Web-based lesson package, Strong & Deadly Futures | • Some facilitators reported that students learned the knowledge that the lesson planned to deliver, while students reported that the substance-related knowledge had been learned and understood. |



XSL•FO
RenderX

| Evidence type and study (author, year) | Data collection method | Data analysis method | Digital mental health service type | Major findings |
|---|---|---|---|---|
| Tighe et al [84], 2020 | Semistructured interview | • Thematic analysis | iBobbly program | • Participants indicated that an app may not be able to provide enough support for vulnerable people who have severe mental health conditions.<br>• The improved mental health literacy helped participants feel distracted from their thoughts and reduced their distress.<br>• Some participants reported that the program has the effect of suicide prevention. |
| Quantitative evidence with comparison to typical mental health services or no intervention groups | | | | |



XSL•FO
RenderX

| Evidence type and study (author, year) | Data collection method | Data analysis method | Digital mental health service type | Major findings |
|---|---|---|---|---|
| Lee et al [72], 2019 | Intervention group: data from the Grog Survey app; Control group: clinical interview | • Triangulating the Finnish method to compare drinking status and average consumption between 2 groups using descriptive statistics.<br>• Agreement between consumption data/daily alcohol consumption and withdrawal tremors was assessed through the Spearman rank correlation coefficient.<br>• Sensitivity and specificity analyses (unspecified) were used to compare how well the app and clinical interview agreed when classifying drinkers' risk. | Grog Survey app | • The app can accurately test participants' drinking status (n=157), with 94.7% agreement with the clinical interview results.<br>• The alcohol consumption level recorded by the app was usually higher than clinical interviews (drinks per drinking occasion: median 17.0, IQR [10.5, 27.9] vs median 15.4, IQR [9.6, 23.2]).<br>• The app tended to categorize more participants at risk. The app identified 76.7% of drinking males and 69.7% of females as being at risk, while the interview identified 74.4% of males and 65.2% of females as being at risk.<br>• The app appeared to be highly sensitive (92.7% - 97.5% sensitivity) in detecting at-risk drinkers, and when using matched and unmatched reference periods, the app could detect 10 and 37 more risky drinkers, respectively, than clinical interviews.<br>• The average daily consumption recorded by the app and by the interview was moderately correlated ($r$=.52). Correlations were greater when only data from matched time periods were used ($r$=.62). Correlations were all statistically significant.<br>• Participants were more likely to report the presence of tremors when using the app than when responding to the interview (17.0% and 7.2%, respectively; $r_s$=0.48).<br>• There was high test-retest reliability for the app ($r_s$=0.81 for both participants and drinkers). |
| Veinovic et al [87], 2023 | | | | |



XSL•FO
RenderX

| Evidence type and study (author, year) | Data collection method | Data analysis method | Digital mental health service type | Major findings |
|---|---|---|---|---|
| | Test results of the face-to-face and telehealth versions of both the MMSE[a] and the KICA-screen[b] tests | • The MMSE and KICA-screen test version total scores were compared using Spearman correlations.<br>• Agreement between test versions was examined using Bland-Altman analyses.<br>• Telephone test scores were prorated to equivalent in-person total scores to examine overall performance discrepancies directly. | Telehealth versions of the MMSE (TMMSE) and KICA-screen (TKICA-screen) | • Despite a reduction in effectiveness, the accuracy of the TMMSE was generally acceptable. The median TMMSE results were slightly lower than the MMSE results (24 vs 28.5, IQR 2 for both). There was a moderate correlation and a reasonable agreement between MMSE versions ($r_s$=0.33; *P*=.20), although the limits of agreement were unacceptably wide (−4.1 and 4.8 point differences).<br>• The TKICA-screen was significantly less effective than the KICA-screen. There was a significant mean difference of 2.17 (95% CI 1.39 - 2.95), with wide limits of agreement (−1.09 and 5.43 point differences) and evidence of proportional bias (B=−1.03, SE=0.32; *P*=.005), with a tendency for poorer agreement at the lower end of the performance. |
| Tighe et al [85], 2017 | Test results from the following psychological instruments: Depressive Symptom Index-Suicidality Subscale (DSI-SS), Patient Health Questionnaire-9 (PHQ-9), Kessler Psychological Distress Scale-10 (K-10), and Barratt Impulsivity Scale-11 (BIS-11) | • Paired *t* test | iBobbly program | |



XSL•FO
RenderX

| Evidence type and study (author, year) | Data collection method | Data analysis method | Digital mental health service type | Major findings |
|---|---|---|---|---|
| | | | | • The effectiveness of the iBobbly app on suicide prevention was insignificant. The pre- and postintervention changes in the DSI-SS score were significant in the iBobbly arm ($t$=2.40, df=58.1; $P$=.02), but these differences were not significant compared with the waitlist arm ($t$=1.05, df=57.8; $P$=.30).<br>• The iBobbly app was effective in reducing depression. The iBobbly arm showed a statistically significant decrease in PHQ-9 scores compared with waitlist controls. The interaction of the iBobbly arm by time was significant ($t$=2.79, df=56.9; $P$=.007; Cohen $d$=0.71, 95% CI 0.17-1.23), reflecting a substantial effect.<br>• The iBobbly app was effective in reducing general psychological distress measured by the K-10. The interaction of the intervention arm by time was significant ($t$=2.44, df=57.5; $P$=.02; Cohen $d$=0.65, 95% CI 0.12-1.17), reflecting a substantial effect.<br>• However, the PHQ-9 and K-10 scores for the control group (comparing baseline and receiving the iBobbly intervention afterwards) were insignificant.<br>• The iBobbly app was not effective in reducing impulsivity, as measured by the BIS-11. The preintervention scores for the waitlist group were significantly lower than those for the iBobbly group ($t$=2.05, df=59.2; $P$=.045). Postintervention means were identical. However, after applying the iBobbly app to the waitlist group, the score reduction was insignificant ($t$=−1.82, df=29.1; $P$=.08). |



XSL•FO
RenderX

| Evidence type and study (author, year) | Data collection method | Data analysis method | Digital mental health service type | Major findings |
|---|---|---|---|---|
| Russell et al [82], 2021 | KICA-screen results obtained from both face-to-face and online tests | • Liner correlation | KICA-screen delivered via telehealth | • The KICA-screen test delivered through digital platforms did not cause a decrease in effectiveness compared to the face-to-face version, owing to good correlation (Pearson $r$=0.851; $P$<.01) and good agreement (intraclass correlation coefficient=0.85; $P$<.01). |
| Titov et al [86], 2019 | Health records measured by the K-10, PHQ-9, and Generalized Anxiety Disorder Scale-7 (GAD-7) | • A generalized estimation equation (GEE) modeling technique was used to examine the significance of changes in symptom measures over time. The main effects of course and Indigenous status were also examined. | MindSpot service | • For First Nations patients, the MindSpot service was effective at causing a significant decrease in the K-10 score ($\chi^2$=5933.1; $P$<.001), PHQ-9 score ($\chi^2$=5938.8; $P$<.001), and GAD-7 score ($\chi^2$=5658.8; $P$<.001).<br>• There were no differences in treatment outcomes between Indigenous and non-Indigenous patients (K-10: $\chi^2$=0.1; $P$=.75; PHQ-9: $\chi^2$=0.5; $P$=.49; GAD-7: $\chi^2$=0.6; $P$=.44).<br>• There was no difference in the outcome between Indigenous patients who chose the general wellbeing course and those who chose the Indigenous wellbeing course (K-10: $\chi^2$=1.2; $P$=.27; PHQ-9: $\chi^2$=0.5; $P$=.48; GAD-7: $\chi^2$=2.9; $P$=.09). |
| Quantitative evidence with no comparison | | | | |



XSL•FO
RenderX

| Evidence type and study (author, year) | Data collection method | Data analysis method | Digital mental health service type | Major findings |
|---|---|---|---|---|
| Dingwall et al [68], 2023 | Comparing participants' scores of psychosocial instruments at baseline and after the 4-week intervention: K-10, Patient Health Questionnaire-2 (PHQ-2), Generalized Anxiety Disorder-short form (GAD-2), Alcohol Use Disorders Identification Test (AUDIT-C), Drug Use Disorders Test (DUDIT), and Parent-rated Strengths and Difficulties Questionnaire (SDQ-parent rated) | • Paired sample 2-tailed *t* tests for psychological test results with within-group effect sizes (Cohen *d*) | AIMhi-Y app | • The app was significantly effective in reducing psychological distress ($t_{29}$=4.63; *P*<.001; Cohen *d*=0.85, 95% CI 2.79-7.21) and depressive symptoms ($t_{29}$=4.09; *P*<.001; Cohen *d*=0.75, 95% CI 0.42-1.25).<br>• Participants presented improvements in reducing anxiety symptoms ($t_{29}$=1.58; *P*=.13; Cohen *d*=0.28, 95% CI −0.10 to 0.77).<br>• The app was ineffective in reducing alcohol use disorder symptoms ($t_{26}$=1.55; *P*=.13; Cohen *d*=0.31, 95% CI −0.06 to 0.43), drug use disorder symptoms ($t_{26}$=1.55; *P*=.13; Cohen *d*=0.31, 95% CI −0.06 to 0.43), and lead dependence ($t_{29}$=0; *P*>.99; Cohen *d*=0, 95% CI −0.29 to 0.29).<br>• Participants' parents reported that the app was ineffective in improving psychological adjustment ($t_3$=2.15; *P*=.12; Cohen *d*=1.07, 95% CI −2.03 to 10.53). |
| Kennedy et al [71], 2021 | The uMARS[c] survey comprises 26 questions and 5-point Likert scales | • Calculate descriptive statistics and mean and SD for continuous variables | Multibehavioral change app "MAMA-EMPOWER" | • Participants (n=16) reported acceptable perceived impact (mean 3.55, SD 1.06). |
| Tighe et al [84], 2020 | Record usage of the app and results from psychological instruments: DSI-SS, K-10, and PHQ-9 | • Regression analyses | iBobbly program | |



XSL•FO
RenderX

| Evidence type and study (author, year) | Data collection method | Data analysis method | Digital mental health service type | Major findings |
|---|---|---|---|---|
| | | | | • The iBobbly app was ineffective for the primary outcome, suicidal ideation, which was measured with the DSI-SS. For every minute spent on the app, DSI-SS scores reduced by 0.013 points ($R^2$=0.29).<br>• The iBobbly app was ineffective in reducing psychological distress that was measured with the K-10. For every minute spent on the app, K-10 scores reduced by 0.007 points ($R^2$=0.35).<br>• The iBobbly app was ineffective in reducing depression that was measured using the PHQ-9. For every minute spent on the app, PHQ-9 scores reduced by 0.001 points ($R^2$=0.268).<br>• According to the authors, the low validity could be caused by the small sample size. |
| Routledge et al [81], 2022 | Paper-and-pen test on classes or an online questionnaire | • Analysis of descriptive data (method unspecified) | Web-based lesson package, Strong & Deadly Futures | • Among First Nations students (n=15), more than half (53.3%) reported that they were likely to use the information and skills taught by the course in their own lives. Nearly half (46.7%) of First Nations students were unsure, and 0% of First Nations students believed they were unlikely to use the information and skills taught by the course.<br>• First Nations students generally thought this program was effective in handling problems related to substance use (73.3% of students supported this idea), stress (66.7%), and peer pressure (80%). |

[a]MMSE: Mini-Mental State Examination.

[b]KICA-screen: Kimberly Indigenous Cognitive Examination-short form.

[c]uMARS: user version of the Mobile Application Rating Scale.



XSL•FO
RenderX

### Targeted Mental Health Conditions

Five studies did not specify the mental health conditions targeted by DMH services [66,67,70,77-80]. Three studies focused on psychological assessment and diagnosis [82,86,87], and 5 studies addressed the management of distress, depression, and anxiety disorders [71,75-85,86]. Additionally, 5 studies examined the management of substance use [71,72,75,76,83], and another 5 studies were related to the management of suicide or self-harm behaviors [67,73,75,84,85]. See Table 2 for detailed information about the settings of the included studies.



XSL•FO
RenderX

**Table .** Settings of the included studies.

| Study (author, year) | Study aim | Participants | Research setting | Digital mental health (DMH) service type | Targeted mental health issue |
|---|---|---|---|---|---|
| Bennett-Levy et al [66], 2017 | Report the barriers and enablers of e-mental health uptake among health providers who provide services to First Nations people | 50 providers of different mental health services for First Nations people | Two locations (Lismore and Tweed Heads) in northern New South Wales | General e-mental health services (did not specify service type) | Unspecified |
| Brown et al [67], 2020 | Understand how mobile apps support suicide prevention gatekeepers in First Nations communities | 12 participants (Indigenous health workers or community members) | A regional city in South West Queensland called Toowoomba | Suicide prevention efforts supported by mobile apps | Suicide and self-harm behaviors |
| Dingwall et al [68], 2023 | Assess the feasibility, acceptability, and use of the Aboriginal and Islander Mental Health Initiative for Youth (AIMhi-Y) app and determine the feasibility of study procedures for future assessments | 30 First Nations young people aged 12 - 25 years | Darwin, Northern Territory (NT) | AIMhi-Y app | Unspecified |
| Dingwall et al [69], 2015 | Develop and determine the acceptability, feasibility, and appropriateness of the Australian Integrated Mental Health Initiative (AIMhi) Stay Strong App for service providers working with First Nations people | 15 service providers, including health professionals, managers, program coordinators, and an Aboriginal elder | Rural and remote primary health care service settings in the NT | AIMhi Stay Strong app | General mental health area (unspecified) |
| Fletcher et al [70], 2017 | Test the acceptability and feasibility of developing the “Stayin’ on Track” website, which offers tailored support and information to young Aboriginal fathers, and adapt and test the mobile phone–based text-messaging and mood-tracker program (SMS4dads program) that provides mental health support to young Aboriginal fathers | 20 young Aboriginal fathers in the assessment of the “SMS4dads” smartphone program; 20 young Aboriginal fathers and an uncertain number of community members in the yarn-up session on the “Stayin’ on Track” website | One regional city and two rural towns in New South Wales | “SMS4dads” smartphone program and “Stayin’ on Track” website | General mental health area (unspecified) |
| Kennedy et al [71], 2021 | Describe the development (stages 1 and 2) and pretest (stage 3) of a prototype multibehavioral change app (MAMA-EMPOWER) | Stage 1: 8 First Nations women; stage 2: 6 First Nations women; stage 3: 16 First Nations women | Newcastle and Tamworth, New South Wales | Multibehavioral change app “MAMA-EMPOWER” | Substance use and psychological distress |



XSL•FO
RenderX

| Study (author, year) | Study aim | Participants | Research setting | Digital mental health (DMH) service type | Targeted mental health issue |
|---|---|---|---|---|---|
| Lee et al [72], 2019 | Validate a survey app (Grog Survey app) that explores the alcohol consumption status among First Nations people | 20 nondrinkers, 40 nondependent drinkers, and 40 dependent drinkers who self-identified as First Nations people and were aged ≥16 years | An Indigenous primary health care service and surrounding community in urban Queensland, and a regional Aboriginal community–controlled health service and a remote Aboriginal community–controlled drug and alcohol day center (a drop-in service) in South Australia | Grog Survey app | Substance use (alcohol consumption) |
| Ma et al [73], 2022 | Understand the expectations of different communities for telehealth crisis support services in Australia | 1300 adults living in Australia, including 2.4% First Nations participants | National setting in Australia | Lifeline Australia | Suicide and self-harm behaviors |
| Perdacher et al [74], 2022 | Determine the feasibility of the Stay Strong app as a digital well-being and mental health tool for use by First Nations people in prison | 27 clients and 10 health practitioners | 3 high-security prisons in Queensland, Australia | Stay Strong App (custodial version) | General mental health area (unspecified) |
| Povey et al [75], 2016 | Explore First Nations community members' experiences of using 2 e-mental health apps and identify the factors that influence acceptability | 9 First Nations adults | Darwin, NT | AIMhi Stay Strong iPad app and the iBobbly suicide prevention app | AIMhi Stay Strong app: general mental health issues and substance use; iBobbly app: psychological distress, depressive symptoms, and suicide prevention |
| Povey et al [76], 2022 | Present an in-depth account of the second phase of participatory design in the development of the AIMhi-Y app | 75 First Nations youth | Across Australia | AIMhi-Y app (under design) | General mental health area (unspecified) |
| Povey et al [77], 2020 | Report the result of the formative stage of the AIMhi-Y app development process that engaged First Nations youth in the co-design of the new culturally informed AIMhi-Y app | 45 First Nations youth | NT, Australia | General e-mental health services (did not specify service type) | General mental health area (unspecified) |
| Puszka et al [78], 2016 | Understand stakeholder perspectives on the requirements for implementing DMH services in regional and remote health services for First Nations people | 32 stakeholders who provide mental health services for First Nations people | NT, Australia | General e-mental health services (did not specify service type) | General mental health area (unspecified) |
| Raphiphatthana et al [79], 2020 | Evaluate the process and effectiveness of the e-mental health program | 66 participants from First Nations primary care organizations | Remote NT communities, Darwin, Alice Springs, and Adelaide | Not specified, but most participants used the Stay Strong app | General mental health area (unspecified) |



XSL•FO
RenderX

| Study (author, year) | Study aim | Participants | Research setting | Digital mental health (DMH) service type | Targeted mental health issue |
|---|---|---|---|---|---|
| Raphiphatthana et al [80], 2020 | Understand service providers' perspectives and experiences of electronic mental health adoption | 57 service providers working with Aboriginal and Torres Strait Islander people who had undergone electronic mental health training workshops | Darwin, Alice Springs, and remote NT communities | General DMH services (but many of the responses were related to the use of the Stay Strong app) | Unspecified |
| Routledge et al [81], 2022 | Assess the acceptability and feasibility of Strong & Deadly Futures, a culturally inclusive alcohol and drug prevention program for First Nations secondary students | 19 First Nations students out of 281 students | Two independent urban schools and two rural state schools | Strong & Deadly Futures, a 6-lesson, web-based package combining online illustrated storylines with interactive classroom activities | Substance use (both alcohol and tobacco) |
| Russell et al [82], 2021 | Examine the utility of using a culturally appropriate dementia screening tool (KICA-screen[a]) in a telehealth setting | 33 medically stable First Nations inpatients/outpatients | Two local rural health care settings | KICA-screen delivered via telehealth | Cognitive assessment |
| Snijder et al [83], 2021 | Describe the development of Strong & Deadly Futures, a web-based substance use education class | 41 First Nations students and 36 non-First Nations students | Four schools in New South Wales | Strong & Deadly Futures, a 6-lesson, web-based package combining online illustrated storylines with interactive classroom activities | Substance use (both alcohol and tobacco) |
| Tighe et al [85], 2017 | Evaluate the effectiveness of a self-help mobile app (iBobbly) targeting suicidal ideation, depression, psychological distress, and impulsivity among Indigenous youth in remote Australia | 61 First Nations Australians aged 18 - 35 years | Remote and very remote communities in the Kimberley region of North Western Australia | iBobbly program | Suicidal ideation, depression, psychological distress, and impulsivity |
| Tighe et al [84], 2020 | Discover pilot usage and acceptability data from the iBobbly suicide prevention app, an app distributed through a randomized controlled trial | 13 First Nations Australians aged 18 - 35 years | Remote and very remote communities in the Kimberley region of North Western Australia | iBobbly program | Suicidal ideation, depression, psychological distress, and impulsivity |
| Titov et al [86], 2019 | Report on Aboriginal and Torres Strait Islander (Indigenous) users of MindSpot, a national service for the remote assessment and treatment of anxiety and depression | 780 First Nations participants | National setting across Australia | MindSpot service, including diagnosis and treatment courses | General mental health area, but this study focused on psychological distress, depression, and anxiety |



XSL•FO
RenderX

| Study (author, year) | Study aim | Participants | Research setting | Digital mental health (DMH) service type | Targeted mental health issue |
|---|---|---|---|---|---|
| Veinovic et al [87], 2023 | Evaluate mental state examinations delivered face-to-face (MMSE[b] and KICA-screen) and as telehealth versions (TMMSE[c] and TKI-CA-screen[d]) among older First Nations people | 20 First Nations people aged 55 - 69 years | Urban and regional New South Wales, Australia | TMMSE and TKICA-screen | Cognitive assessment |

[a]KICA-screen: Kimberly Indigenous Cognitive Examination-short form.

[b]MMSE: Mini-Mental State Examination.

[c]TMMSE: telehealth version of the MMSE.

[d]TKICA-screen: telehealth version of the KICA-screen.

## Quality Assessment Results

According to the results from the Aboriginal and Torres Strait Islander Quality Appraisal Tool, the included studies presented an acceptable connection to First Nations values and principles, with at least 16 studies having "Yes" or "Partially" (provided sufficient information) answers to 11 out of 14 questions. However, most studies (n=14-18) did not properly negotiate a formal agreement with First Nations people to access and protect their knowledge. Moreover, some studies were deemed less culturally engaging for First Nations people, such as a study that researched DMH services designed for the general population [73], and some studies applied quantitative methods that resulted in lower participation in First Nations communities [82,86,87].

In terms of the results from the MMAT, all studies were considered to have high-quality research designs. For example, all qualitative studies, except for 1 study [71], completely matched all MMAT criteria. Moreover, all included studies had at least 82% of the criteria marked as "Yes."

## Effectiveness

Twelve studies [68,71,72,74,75,80-82,84-87] reported the effectiveness of DMH services in managing mental health conditions among First Nations people in Australia (Table 1). Their results are summarized in the following paragraphs based on their study designs.

### *Qualitative Evidence*

Six studies used a qualitative study design to explore the perceived effectiveness of DMH services in managing mental health conditions among First Nations people [68,74,75,80,81,85]. Three studies [68,74,75] assessed the youth version, custodial version, or standard version of the Aboriginal and Islander Mental Health Initiative (AIMhi) Stay Strong app for First Nations people. The AIMhi Stay Strong app helped participants achieve mental health improvements, guided positive thinking, and supported participants in overcoming difficulties [68,74,75]. Participants perceived the AIMhi Stay Strong app and general DMH services to be effective for improving their mental health outcomes [80]. Moreover, "Strong and Deadly Futures" online courses were examined among First Nations people to prevent alcohol and tobacco use and reduce stress and peer pressure [81]. Both clients and practitioners reported that the app enhanced clients' self-insight, self-reflection, positive thinking, confidence, and overall self-perception. However, in the study by Povey et al [75], while both the AIMhi Stay Strong and iBobbly apps were viewed as helpful for managing minor mental health difficulties, participants considered them less suitable for addressing severe conditions or preventing suicide. Participants agreed that the efficacy of the iBobbly app was limited for "vulnerable people with strong emotions" [84].

### *Quantitative Evidence*

Eight studies reported quantitatively measured effectiveness [68,71,72,82,84-87], with mental health outcomes measured through self-reported instruments and scales.

### *Quantitative Evidence Without Comparison*

Four studies reported the effectiveness of DMH services in managing mental health conditions among First Nations people without comparing them to typical services or other control groups [68,71,81,84]. Kennedy et al [71] found that the multibehavioral change app "MAMA-EMPOWER" was effective in managing substance use and psychological distress and was acceptable (average score of 3.55 on a 5-point Likert scale) for First Nations mothers. Meanwhile, Routledge et al [81] found that Strong & Deadly Futures (a web-based lesson package aiming to improve mental well-being and substance use prevention) was effective in 53.3% of First Nations students, who were likely to use the knowledge they obtained from the course, and no participants reported that they were unlikely to apply that knowledge. Moreover, 73.3%, 66.7%, and 80.0% of First Nations students thought that this program was helpful in managing substance use, stress, and peer pressure, respectively. However, Dingwall et al [68] found mixed therapeutic outcomes for the AIMhi-Y general mental health support app among First Nations youth. More specifically, the app was found to be effective in reducing psychological distress and depressive symptoms (Cohen $d$=0.85 and 0.75, respectively) and reducing anxiety symptoms (without statistical significance), but was not effective in managing substance use disorders or improving psychological adjustment. Furthermore, Tighe et al [84] found that the iBobbly app was not effective in managing suicidal ideation and reducing psychological distress or depression.



XSL•FO
RenderX

However, all indicators presented a trend of improvement. See Table 1 and Multimedia Appendix 3 for detailed information.

### *Quantitative Evidence With Comparison*

Five studies examined the effectiveness of DMH services in managing [85,86] and assessing [72,82,87] mental health conditions by comparing the effectiveness of DMH services with that of typical mental health services [72,82,87] or waitlist control groups [85], or comparing the effectiveness of DMH services between First Nations people and non-First Nations people [86].

The Grog Survey app [72], which was developed for assessing substance abuse among First Nations people, was found to be effective in detecting the alcohol consumption level (94.7% overall agreement between the app and clinical interviews). The app was also found to be more sensitive than a clinical interview for the number of drinks, the greatest number of drinks, the median of reported drinking, and the daily consumption of standard drinks per drinking occasion (mean difference=1.6, 0.8, 3, and 0.4, respectively). Higher sensitivity was found among those classified as risky drinkers (2.3% and 4.5% more detection among males and females, respectively), as well as in the reporting (17.0% vs 7.2%) and prediction ($P$=.02 vs $P$=.44) of tremors. See Table 1 and Multimedia Appendix 3 for detailed information.

A previous study compared the telehealth versions of the Mini-Mental State Examination (TMMSE) and Kimberly Indigenous Cognitive Examination-short form (TKICA-screen), which were used for cognitive screening, with their typical versions [87]. The study found that the median TMMSE scores were slightly lower than the Mini-Mental State Examination (MMSE) scores (median difference=4.5), and the agreement measured by correlation was moderate ($r_s$=0.33; $P$=.20) but with unacceptable limits of agreement (the effectiveness may not significantly reduce in the TMMSE). On the other hand, the TKICA-screen showed a significant effectiveness loss compared with the Kimberly Indigenous Cognitive Examination-short form (KICA-screen). For instance, there was a significant mean difference of 2.17, with a tendency for poorer agreement at the lower end of the performance. See Table 1 and Multimedia Appendix 3 for evidence of proportional bias. Russel et al [82] found that the KICA-screen delivered through digital platforms was as effective as the face-to-face version, with a statistically significant correlation (Pearson $r$=0.851; $P$<.01) and agreement (intraclass correlation coefficient=0.85; $P$<.01) between the versions.

Titov et al [86] found that the MindSpot service, aiming to provide general mental health support, was effective in managing psychological distress, depression, and anxiety, with statistically significant decreases in the Kessler Psychological Distress Scale-10, Patient Health Questionnaire-9, and Generalized Anxiety Disorder Scale-7 scores ($\chi^2$=5933.1, 5938.8, and 5658.8, respectively). As a DMH service designed for all Australians, there were no statistically significant differences in treatment outcomes between First Nations patients and non-First Nations patients or between First Nations patients who chose the general well-being course and those who chose the First Nations well-being course.

Tighe et al [85] evaluated the iBobbly app that was designed for suicide prevention. The results suggested that there were no statistically significant differences between the iBobbly app group and the waitlist control group in suicide prevention ($t_{58.1}$=2.40; $P$>.05). However, it was effective in managing depression when compared with a waitlist control group and preintervention baseline scores ($P$<.05). Similarly, the DMH service was found to be effective in managing psychological distress ($t_{57.5}$=2.44). See Table 1 and Multimedia Appendix 3 for $P$ values and effect sizes. The iBobbly app was found to not be effective in managing impulsivity.

## Determinants of DMH Service Implementation

### *Overview*

Overall, this review identified the determinants of facilitators and barriers that influence the implementation of DMH services. See Table 3 for the extracted themes.

Table 3 presents 9 themes and 26 subthemes that describe the facilitators and barriers influencing the implementation of DMH services for First Nations people. There were 4 subthemes under *organizational factors*, 4 subthemes under *cultural appropriateness*, 6 subthemes under *accessibility*, 1 subtheme under *engagement*, 2 subthemes under *coverage*, 6 subthemes under *acceptability*, 1 subtheme under *integration of DMH services with existing context*, 4 subthemes under *digital literacy*, and 1 subtheme under *efficiency of DMH services*.



XSL•FO
RenderX

**Table .** Themes and subthemes of the identified studies.

| Theme and subtheme | Supporting evidence |
|---|---|
| Theme 1: Organizational factors within DMH[a] service providers | |
| Supportive structure and managers within DMH service provider organizations facilitate DMH service uptake | • Supportive structures or working culture within organizations facilitated resource allocation or staff training for DMH service uptake [78-80].<br>• Enthusiastic managers provided direct support, which increased the willingness of staff and the organization itself for DMH service uptake [66,80]. |
| Policy and structural obstacles and a lack of resources in DMH service provider organizations create barriers for DMH service uptake | • Procedural and administrative obstacles and obstructive IT policies limited the access of staff to technologies and DMH services and created failures in digital resource allocation [66,67,79,80].<br>• High workloads within the workplace and a lack of digital resources resulted in a lack of DMH training for DMH service providers [66,79,80]. |
| Theme 2: Cultural appropriateness of DMH services | |
| DMH services could be culturally appropriate owing to cultural safety designs and culturally relevant content in DMH services and empowerment, which can increase the willingness of First Nations people to use DMH services | • Overall cultural safety reported by participants [74].<br>• DMH services empowered clients to record their own information and interact with service providers [78,80].<br>• Cultural safety designs were applied in DMH services, which created a safe environment or directly increased the acceptability and interest of clients [69,70,75,76,83,84]. |
| Lack of cultural appropriateness creates barriers for First Nations users to overcome existing negative perceptions and accept DMH services | • Negative perception toward the health system created barriers for First Nations people to use DMH services [67].<br>• Problematic use of English within DMH services and lack of First Nations languages may cause difficulties in using DMH services [75,83]. |
| Theme 3: Accessibility of First Nations people to DMH services | |
| First Nations people have mixed perceptions of their accessibility to DMH services | • First Nations clients generally appraised the high accessibility of DMH services, such as reporting that the DMH services they received were accessible [83,86].<br>• However, general accessibility-related concerns, such as a lack of services and resources, and long response times, were also expressed [67]. |
| Portability and flexibility of DMH services, as well as existing digital infrastructure coverage among First Nations communities, could facilitate the implementation of DMH services | • DMH services were available on multiple digital platforms [75].<br>• A high coverage of digital infrastructure among communities was presented, which facilitated the uptake of DMH services [67,77].<br>• DMH services can be operated on portable devices, which can create unique advantages in accessibility [78,84]. |



XSL•FO
RenderX

| Theme and subtheme | Supporting evidence |
|---|---|
| Many First Nations people report limited access to appropriate digital infrastructure, resulting in low accessibility to DMH services | • In some First Nations communities, a lack of digital infrastructure and devices resulted in difficulties in the implementation of DMH services [69,77-80].<br>• Even in special environments, such as prisons, access to digital devices and some specific functions of DMH services, such as access to images, was limited [74].<br>• In remote areas, a lack of digital infrastructure created barriers for First Nations residents to implement DMH services [75,78]. |
| Free DMH apps are highly accessible, while those with download fees limit client access | • Apps with high accessibility were preferred, such as free-to-download apps that can operate on different platforms [76].<br>• The cost of e-mental health apps negatively influenced their accessibility [75]. |
| DMH services that do not have First Nations language versions create accessibility barriers for First Nations people whose first language is not English | • A lack of DMH services in First Nations languages created accessibility barriers [69,75,78]. |
| Theme 4: Ability of DMH services to promote engagement | |
| DMH services can promote engagement between clients and service providers by breaking the power imbalance and then creating facilitating effects | • DMH services could help establish positive relationships between health care providers and clients, eventually creating positive experiences for clients or increasing First Nations clients' willingness to use these services [69,74,78]. |
| Theme 5: Ability of DMH services to cover a wide range of settings and First Nations population groups | |
| DMH services could cover different types of clients, topics, and reference periods | • High applicability across different age groups and mental health conditions [69].<br>• DMH services could cover sensitive topics due to their nature of indirect engagement [69,78].<br>• Compared to typical face-to-face mental health services, DMH services could cover longer reference periods on substance use topics [72]. |
| DMH services fail to cover some severe mental health conditions, which is considered a barrier | • DMH services may not be able to cover severe mental health conditions, such as suicide [75,78]. |
| Theme 6: Acceptability of First Nations people to DMH services | |



XSL•FO
RenderX

| Theme and subtheme | Supporting evidence |
|---|---|
| Acceptability of First Nations people to DMH services is perceived as high | • Satisfaction and optimism regarding DMH services and willingness to use DMH services to seek help were presented by First Nations clients, indicating high acceptability [68,69,71,74,75,77,81,84,86].<br>• First Nations people also expressed their idea of not expecting to receive help from DMH services [73]. |
| Good visual design can facilitate participants' acceptance of DMH services, while visual design problems are considered barriers | • Clients reported that the DMH services they received had an attractive or engaging visual design, which increased their acceptability of DMH services [68,69,71,74,78].<br>• Some visual design problems also raised the attention of clients, such as high text density [70]. |
| The unique nature of DMH services has high acceptability among First Nations people | • First Nations clients preferred the web-based, client-led, and clinician-supported nature of DMH services [74,75,83]. |
| Attractive and useful content included in DMH services facilitates First Nations people to accept DMH services | • First Nations clients perceived that mental health resources in DMH services were generally interactive and useful [71,80].<br>• Embedded media forms, such as games, videos, and animations, were especially preferred by First Nations clients [68,75].<br>• Tailored content for First Nations students was welcomed [83]. |
| Inadequate, repetitive, and less engaging content included in DMH services is considered a barrier | • Inadequate and repetitive content in DMH services reduced the acceptability of DMH services [68,71,84]. |
| The unique functions of DMH services increase their acceptability among First Nations clients | • The unique functions of DMH services, such as text messages sent as reminders or data capture functions, were acceptable for First Nations clients and prompted clients to use those services [68,69,74,78]. |
| Some functions are not attractive enough or are difficult to use, resulting in lower acceptability of DMH services | • However, some functions of DMH services were ignored, could hardly be used, or could not fit in special contexts, such as in prison [68,71,74,75]. |
| DMH services with data safety have high acceptability, but the concern about data safety is perceived as a barrier | • DMH service apps with high privacy and data safety were preferred [76].<br>• Some service providers and clients also considered data unsafety as a potential risk and barrier [67,78]. |
| Theme 7: Ability of DMH services to integrate with existing contexts | |
| Some DMH services are hard to integrate with existing contexts | • Some service providers felt that it was challenging to integrate DMH services into their routine care because DMH services may not easily fit group clients, pre-existing practices, and local community status [79,80]. |
| Theme 8: Digital literacy of clients and service providers | |



XSL•FO
RenderX

| Theme and subtheme | Supporting evidence |
|---|---|
| Clients and service providers who have high digital literacy facilitate the uptake of DMH services | • Young clients presented especially high digital literacy and willingness to receive DMH services [67,80].<br>• Based on their own enthusiasm for using DMH services or specific DMH training, some service providers obtained high DMH literacy [66,67]. |
| Clients who do not have high digital literacy may face challenges when operating DMH services | • Some clients reported that the DMH app is difficult to operate [70,75,78].<br>• Relatively lower digital literacy resulted in lower awareness of DMH services [68,75,80]. |
| Service providers who do not have high digital literacy may face challenges when operating or providing DMH services | • Low digital literacy of service providers led to negative perceptions, including overestimating the complexity of technology or thinking DMH services will distract them from their current job [66,67,78].<br>• Inadequate digital literacy among service providers created challenges in maintaining and using DMH services [67,79,80].<br>• Overall, the digital literacy level among service providers was highly heterogeneous [78,79]. |
| Theme 9: Efficiency of DMH services | |
| DMH services have high efficiency and are considered time-saving | • Service providers believed that DMH services could reduce the time cost of recording client notes and waiting for data transmission, and these had advantages for data recording and management compared to typical paper format services [78,80]. |

[a]DMH: digital mental health.

### *Organizational Factors*

Organizational factors related to the administration and management of the organization (usually service providers) could either facilitate or create barriers to implementing DMH services.

Four studies asked stakeholders and service providers to generally comment on DMH services [66,78-80]. The supportive structure within the organizations was a significant facilitator, which included the regular supervision and review of DMH service utilization [78]; being supportive of allocating resources, time, and training [79]; and building the work culture that welcomes DMH services within the workforce [80]. Direct support from managers was also considered to facilitate the uptake of DMH services [80] by increasing the opportunity and interest of staff in adopting DMH services [66].

Four studies asked mental health service providers about their general perspectives on DMH services [66,79,80] and on using DMH services in suicide prevention [67]. Service providers reported that organizational factors created barriers for the uptake of DMH services. These included a lack of organizational support [80], administrative obstacles, obstructive IT policies within the workplace [66,79], work restrictions [67], a lack of training resources within organizations [79], and time allocated to cope with the extra workload brought on by the implementation of DMH services [66,79,80].

### *Cultural Appropriateness*

Cultural appropriateness was perceived as a critical determinant of the uptake of DMH services within First Nations people in Australia, with high cultural appropriateness facilitating uptake and low cultural appropriateness impeding uptake.

Ten studies reported high culturally appropriate DMH services [69-71,74-76,78,81,83,84]. The users of the AIMhi Stay Strong app [69] and its custodial version [74] reported satisfaction with cultural safety. Similarly, participants in programs, such as the “SMS4dads” mobile app program, “Stayin’ on Track” website program [70], AIMhi-Y app program [76], and iBobbly suicide prevention program [84], valued the culturally appropriate visual and content design [76] as well as the private space these tools provide to reduce shame and stigma [75,84]. Stakeholders also highlighted the empowering nature of DMH services that allowed users to record personal information and engage with culturally meaningful content [71,78].

The “Strong & Deadly Futures” web-based substance use prevention program, though designed for the general population, was also praised for its cultural appropriateness owing to its inclusion of First Nations characters and an equitable learning



XSL•FO
RenderX

environment for both First Nations and non-First Nations students [81,83].

Three studies reported a lack of cultural safety within DMH services [67,75,83]. Some First Nations health workers perceived digital suicide prevention tools as culturally inappropriate, reflecting broader concerns about cultural unsafety in the mental health system [67]. Users of the AIMhi Stay Strong app and the iBobbly suicide prevention app expressed uncertainty about whether these tools could address trauma linked to colonization [75], and "Strong & Deadly Futures" participants provided mixed responses for the use of "Aboriginal English" [83].

### *Accessibility*

Accessibility was another key determinant of DMH service implementation among First Nations people. Two studies generally evaluated accessibility [83,86], noting that MindSpot could deliver care when local face-to-face mental health services were unavailable [86], and Strong & Deadly Futures was accessible to users with varying abilities [83]. However, concerns remained about poor communication systems, limited resources, slow response times, and service gaps in digital suicide prevention tools [67].

Five studies reported that improved digital infrastructure supported access to DMH services [67,75,77,78,84]. Apps, such as AIMhi and iBobbly, were considered accessible due to their compatibility with multiple devices and their portability [67,75,77,78,84], which enabled use across locations and levels of remoteness. Nevertheless, limited availability of devices and internet connectivity reduced access [78-80], particularly in remote communities and custodial settings [74,75].

Cost was another barrier, with participants preferring free-to-download apps, and the fees for the AIMhi Stay Strong and iBobbly apps were seen as deterrents [75,76].

Finally, language barriers restricted access, as most apps lacked First Nations language options, making it difficult for some users to understand key terms such as "resilience" [69,75,78].

### *Integration*

The difficulty in integrating DMH services with existing practices was generally perceived as a barrier in the included studies. More specifically, some service providers in 2 studies commented on applying DMH services in general mental health areas [79,80] and reported difficulties in integrating DMH services into usual practices when facing non-one-to-one settings, such as facing a group of clients [79], or the need to fit existing practices and the local community status [80].

### *Engagement (Between Service Providers and Clients)*

Three studies highlighted that DMH services promoting engagement between clients and service providers were well received in general mental health care [69,74,78]. The AIMhi app was reported by health providers to reduce barriers, such as power imbalance, and foster positive relationships between health care providers and clients [69]. Moreover, stakeholders expanded this result to other DMH tools that could provide indirect engagements when solving sensitive issues and supported the building of positive relationships [78]. Improved rapport was also observed in custodial settings through the custodial version of the Stay Strong app [74].

### *Coverage*

Coverage refers to the ability of DMH services to cover different types of clients and various mental health conditions. Three studies found that broader coverage facilitated implementation [69,72,78]. The AIMhi app was reported to reach clients across age groups and symptom profiles, including those who were otherwise difficult to engage [69]. Moreover, clients of the AIMhi app [69] and service providers providing general opinions on DMH services [78] noted that DMH services could address sensitive mental health topics. Furthermore, the Grog Survey app, which aimed to investigate the alcohol dependence status, enabled coverage of longer reporting periods than typical clinical interviews [72].

However, 2 studies reported that clients and service providers felt DMH services, including the AIMhi and iBobbly apps, may be unsuitable for addressing severe mental health conditions such as suicide [75,78].

### *Acceptability*

Acceptability refers to the extent to which DMH services are acceptable to First Nations Australians. Ten studies mentioned overall positive acceptability [68,69,71,73-75,77,81,84,86]. Except for Lifeline Australia, which was rated low due to participants' limited expectations of receiving suicide prevention support [73], all other services, including the AIMhi Stay Strong app and its custodial and youth versions [68,69,74], "MAMA-EMPOWER" [71], AIMhi, iBobbly [75,84], "Strong & Deadly Futures" [81], and MindSpot [86], were generally well received, showing high satisfaction, usability, and acceptability [77].

The nature of DMH services was reported as acceptable by 3 studies [74,75,83]. The custodial version of the AIMhi app was appraised for its client-led nature [74], and the AIMhi app and iBobbly suicide prevention app were praised for their clinician-supported nature that could overcome mental health literacy barriers [75]. The web-based nature of "Strong & Deadly Futures" was preferred by students of web lessons [83].

Two studies emphasized relevance and noted that MAMA-EMPOWER was meaningful to users facing relevant issues, although substance use content was viewed as less relevant by nonusers or younger students [71,81].

Four studies discussed digital functions, identifying useful features, such as text messaging in AIMhi [68], service consistency functions in AIMhi [69], goal-setting in the custodial Stay Strong app [74], and data capture features [78]. However, some functions were ineffective or difficult to use, such as the "stories" page in AIMhi-Y [68], and had limited avatar customization [74] and hard-to-use designs [71,75].

Five studies described content as a major contributor to acceptability [68,71,75,80,83]. Participants valued engaging and informative material, including media components (games, videos, and graphics) and culturally tailored content for First



XSL•FO
RenderX

Nations users [83]. Conversely, inadequate or repetitive content reduced engagement [68,71,84].

Six studies addressed visual design [68-71,74,78]. Most participants praised the cultural appropriateness, color, and user-friendly layout of AIMhi and MAMA-EMPOWER, while some found SMS4dads and Stayin' on Track overly text-heavy [70]

Finally, 3 studies discussed data safety and privacy as concerns [67,76,78]. Youth participants preferred DMH tools with secure data handling [76], but general apprehension about data security, particularly in suicide prevention contexts, was reported as a barrier [67,78].

### *Digital Literacy*

Digital literacy was identified as a key determinant of the implementation of DMH services. Three studies reported that high digital literacy facilitated the use and acceptability of DMH services [66,67,80]. Young First Nations clients with higher digital literacy demonstrated greater acceptance of DMH tools [67,80], and some suicide prevention gatekeepers preferred digital platforms, such as Facebook, for service delivery [67]. Training also improved providers' digital literacy, enhancing DMH adoption [66].

However, low digital literacy hindered engagement for both clients [68,70,75,78,80] and service providers [66,67,78-80]. Clients reported difficulty operating apps, such as SMS4dads, Stayin' on Track, AIMhi, and iBobbly [70,75,78], with some forgetting or being unaware of DMH tools [75]. Among providers, digital competency varied widely, with some describing themselves as "less technologically competent" and struggling with basic functions like downloading apps [67,79,80].

### *Efficiency*

Efficiency was another determinant of the implementation of DMH services. Two studies [78,80] reported that service providers perceived DMH services as improving efficiency. More specifically, DMH services could reduce the time cost of recording client notes and waiting for data transmission [78], offering advantages in data recording and management compared with traditional paper-based systems [80].

## Discussion

### Principal Findings

This is the first review to synthesize evidence on the effectiveness of DMH services and the determinants of the facilitators and barriers of DMH services for First Nations people in Australia. Our review found that DMH services were effective in assessing mental health conditions, monitoring mental health status, and delivering mental health education, and had the potential to improve mental health conditions among First Nations Australians. However, the effectiveness of these services for severe mental health conditions remains limited. Additionally, this review identified determinants that could facilitate or inhibit the uptake of DMH services among First Nations people in Australia.

### Effectiveness of DMH Services

The effectiveness of DMH services for First Nations Australians varied across different mental health outcomes and intervention types. Consistent with a previous government report [88] and similar studies [89-91], qualitative studies included in this review suggested that DMH interventions can support mental health improvement, increase self-reflection, and enhance self-confidence and empowerment [68,74,75,80,81,84].

Considering that most included DMH services applied culturally adapted designs, these findings align with broader evidence that culturally adapted DMH tools can provide valuable support for the mental health of First Nations people [3,10,92]. However, concerns remain about the adequacy of digital interventions for severe mental health conditions, particularly suicide prevention, where some participants expressed doubts about their effectiveness beyond mild to moderate distress [75,84,85]. This highlights the ongoing challenge of delivering digital suicide prevention interventions effectively in First Nations communities, as also noted in previous research [55].

Moreover, evidence supports the effectiveness of DMH interventions for substance use. The Grog Survey app accurately identified at-risk drinkers with high sensitivity and reliability [72], and it was as effective as the web-based lesson package Strong & Deadly Futures in preventing substance use [81]. However, the AIMhi-Y app was ineffective in reducing alcohol and drug use symptoms [68], indicating that while DMH tools could be valuable for monitoring and performing early intervention, they might not have sufficient therapeutic impact for substance use disorders without additional face-to-face support. This aligns with previous research on First Nations people in the United States, who had insignificant therapeutic outcomes [93]. The use of DMH interventions for cognitive assessment has also yielded mixed results [82,87]. These findings highlight the need for careful adaptation of cognitive assessments for telehealth delivery, as the telehealth version of the test reduced the visual cues of the test [87] and resulted in insignificant results, but the same test that was delivered through video conferencing obtained acceptable effectiveness compared with typical in-person services [82]. Finally, our engagement and user perception findings indicated that DMH education and behavioral change interventions can be beneficial but require further refinement. For instance, the Strong & Deadly Futures program was generally acceptable to First Nations students owing to the perceived effectiveness in preventing substance use [81]. However, a substantial proportion of students remained unsure about applying the information in their own lives. This suggests that while digital education programs can improve mental health literacy, their impact on long-term behavioral change may be limited without additional reinforcement, as noted in previous research [94].

This review found that the effectiveness of DMH services for First Nations people in Australia is in line with that for First Nations people from other countries. For example, the effectiveness of DMH services in Australia was found in international evidence, such as a general mental health program that successfully caused positive behavioral change for First Nations people in Canada [95] and a mental health screening



XSL•FO
RenderX

program in New Zealand [96]. This indicates that the experience and knowledge generated from First Nations people can be generalized. Heterogeneity was present in some areas. For example, a program in the United States achieved statistical significance in treating substance use issues for First Nations people [93], while a program with a similar target only obtained insignificant results in Australia [68]. Considering that the situation and historical factors faced by First Nations people in Australia are unique, care should be taken when generalizing the results from this review.

### Determinants of the Uptake of DMH Services Among First Nations People in Australia

This review highlights that the implementation of DMH services for First Nations people is shaped by interconnected structural, cultural, and individual factors. Organizational readiness, including leadership support, workforce capacity, and digital infrastructure, remains foundational. Leadership drives the strategic vision, fosters innovation, and ensures resource allocation, which are essential for successful implementation [97,98]. Strong leadership also facilitates stakeholder engagement and addresses resistance to change, which are key factors in digital health uptake [99]. Organizational support further enhances workforce readiness and policy development, sustaining long-term adoption [100]. Without these elements, DMH initiatives risk poor adoption and limited impact [66,67,79,80]. Additionally, cultural appropriateness emerged as a defining determinant of uptake. Culturally safe, co-designed tools enhanced engagement and empowerment, whereas culturally unsafe or linguistically limited designs hindered trust and acceptability. Culturally relevant app designs incorporating First Nations art and metaphors were found to enhance engagement [69,70,75,76,83,84]. This echoes prior research emphasizing the importance of embedding cultural elements to ensure relevance and acceptance within First Nations communities [101].

Accessibility, digital literacy, and efficiency collectively reflect the broader digital divide that continues to shape health equity in Australia. First, inconsistent digital literacy among both clients and practitioners has been previously identified as a key obstacle to effective use of DMH services [66-68,70,75,78-80]. Delivering technology training and ongoing support could improve digital literacy and confidence in using DMH services [102,103], ultimately enhancing their uptake and efficacy among First Nations communities. Second, while improved infrastructure and targeted training enhanced access and usability in some First Nations communities, persistent disparities in internet connectivity, device availability, and digital skills constrained the scalability of DMH interventions, particularly in remote and custodial settings [104]. Despite evidence showing common ownership of digital infrastructure in some First Nations communities [67,77], limited infrastructure and technology access remain significant challenges, particularly in remote communities where access to tablets, Wi-Fi, and reliable technology is often restricted [67,69,75,77-80]. It has been reported that First Nations Australians, especially those living in remote areas, are some of the most digitally excluded populations [105]. This finding highlights the urgent need for targeted investments in digital infrastructure to bridge the technology gap and ensure equitable access to DMH services. To address these challenges, it could be helpful to expand community Wi-Fi programs, provide subsidized digital devices, and implement mobile-based DMH solutions that require minimal bandwidth [88]. The Mob Link initiative by the Institute for Urban Indigenous Health is a strong example [106], as it provides culturally responsive digital health support, including assistance with telehealth and digital access, helping to reduce barriers and improve engagement with health care services. Without sustained investment in digital inclusion and community-based training, DMH services risk reproducing existing inequities rather than closing service gaps [107].

At the service level, DMH tools enhance engagement and communication between clients and providers, supporting self-reflection and rapport. However, limited integration with existing care models, concerns about data privacy, and uncertainty about applicability to severe mental health conditions reduce trust and engagement with DMH services and signal the need for clearer clinical pathways and blended-care frameworks. For instance, privacy concerns and the fear of data misuse have been constantly identified as significant deterrents to the uptake of DMH services [67,78]. It is imperative to address the concerns of First Nations people regarding data sovereignty and the potential misuse of personal health information [108], so that DMH services can be designed with robust privacy protection and transparent data governance frameworks. Furthermore, streamlined data management and integration may be helpful for integrating DMH services and reducing administrative burden, enabling system-level adoption of DMH services for First Nations Australians.

### Heterogeneity of DMH Service Types

This review included various DMH service types, which caused differentiation in determinants. For example, the telehealth version of an included cognitive screening tool showed lower effectiveness due to the absence of visual cues in the test [87], while the same test with visual cues obtained acceptable effectiveness [82]. A possible explanation for this difference is that the latter test was delivered through video conferencing and involved a complete version of the assessment containing images and visual cues. Similarly, the visual design and multimedia content of DMH apps [68,69,71,74,75,78,80,83] and the various digital functions in DMH apps [68,69,74,78] were praised for increasing the acceptability of DMH services among First Nations people. However, no similar facilitators were reported by the users of telehealth DMH services, hinting at the potential benefits of designing app- or web-based DMH services compared with telehealth DMH services. Despite such diversities, this review aimed to synthesize all evidence representing technology-assisted mental health services. Therefore, it has provided an integrated understanding of how digital approaches have been applied to support mental health among First Nations people.

### Limitations

We acknowledge that this review has several limitations. First, the included studies were highly heterogeneous in terms of intervention types, study designs, and outcome measures. This limited our ability to draw precise or generalizable conclusions



XSL•FO
RenderX

about overall effectiveness. Second, as our search was restricted to academic databases, relevant gray literature may have been missed, potentially contributing to publication bias. Third, many included studies had a moderate to high risk of bias, with small sample sizes and self-reported outcomes, which may reduce the certainty of the evidence.

## Conclusion

This review highlights the effectiveness of DMH services in assessing, monitoring, and managing mental health conditions for First Nations Australians. However, information on their impact on severe conditions, including suicide, remains limited. Digital health apps showed promising results in improving general mental health support. Digital health apps and online courses were also perceived to be effective in preventing and assessing substance use, but did not have a significant treatment effect on managing it. For screening mental health conditions, the digital health versions of instruments were comparable to face-to-face versions, while the telehealth versions showed less effectiveness than both the digital health and face-to-face versions, indicating that effectiveness may vary across different types of DMH services.

Long-term funding in DMH services through culturally responsive community services (eg, Institute for Urban Indigenous Health) from the government or organizations should be beneficial, especially for First Nations people facing barriers to accessing in-person mental health services (eg, those who reside in remote areas) and for DMH services that are supported by evidence.

This study also identified key facilitators (eg, strong leadership, culturally relevant designs, clinician-supported tools, and community) and major barriers (eg, digital exclusion, low digital literacy, and privacy concerns). These findings highlight the need for targeted infrastructure investment that can reduce the digital divide for First Nations people, especially in regional and remote communities; digital training for both mental health care providers and First Nations clients; and transparent data governance. However, the evidence should be interpreted with caution, given the heterogeneity of study designs and outcomes, potential publication bias, and moderate to high risk of bias across the included studies. Future research should focus on co-designing DMH services with First Nations communities, rigorously evaluating their effectiveness, and integrating them into existing health care models to ensure accessibility, cultural relevance, and long-term impact.

## Acknowledgments

Artificial intelligence (ChatGPT, OpenAI) was used solely to refine grammar and improve expression clarity. All authors reviewed and approved all content after artificial intelligence–assisted editing.

## Funding

No funding is involved.

## Data Availability

All data generated from this study are publicly available.

## Authors' Contributions

Conceptualization: SZ, ACS, SD, XZ
Investigation: SZ, SD, XZ
Methodology: SZ, AG, SD, XZ
Project administration: ACS
Supervision: SD, XZ
Writing – original draft: SZ
Writing – review & editing: AG, ACS, SD, XZ

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed inclusion and exclusion criteria, and PICO (population, intervention, comparison, outcome) protocol.
[DOCX File, 19 KB - i-jmr_v15i1e80386_app1.docx ]

Multimedia Appendix 2
Details of the search strategy.
[DOCX File, 19 KB - i-jmr_v15i1e80386_app2.docx ]

Multimedia Appendix 3



XSL•FO
RenderX

Details of the extracted data.
[XLSX File, 34 KB - i-jmr_v15i1e80386_app3.xlsx ]

Multimedia Appendix 4
Quality assessment results (Mixed Methods Appraisal Tool).
[XLSX File, 14 KB - i-jmr_v15i1e80386_app4.xlsx ]

Multimedia Appendix 5
Quality assessment results (Aboriginal and Torres Strait Islander Quality Appraisal Tool).
[XLSX File, 14 KB - i-jmr_v15i1e80386_app5.xlsx ]

Checklist 1
PRISMA checklist.
[DOCX File, 26 KB - i-jmr_v15i1e80386_app6.docx ]

## References


1. Butler TL, Anderson K, Garvey G, et al. Aboriginal and Torres Strait Islander people's domains of wellbeing: a comprehensive literature review. Soc Sci Med 2019 Jul;233:138-157. [doi: 10.1016/j.socscimed.2019.06.004]
2. Ralph S, Ryan K. Addressing the mental health gap in working with Indigenous youth: some considerations for non‐Indigenous psychologists working with Indigenous youth. Aust Psychol 2017 Aug 1;52(4):288-298. [doi: 10.1111/ap.12287]
3. Farah Nasir B, Brennan-Olsen S, Gill NS, et al. A community-led design for an Indigenous model of mental health care for Indigenous people with depressive disorders. Aust N Z J Public Health 2021 Aug;45(4):330-337. [doi: 10.1111/1753-6405.13115] [Medline: 34097328]
4. Parker R. Australia's Aboriginal population and mental health. J Nerv Ment Dis 2010 Jan;198(1):3-7. [doi: 10.1097/NMD.0b013e3181c7e7bc] [Medline: 20061862]
5. Nasir BF, Toombs MR, Kondalsamy-Chennakesavan S, et al. Common mental disorders among Indigenous people living in regional, remote and metropolitan Australia: a cross-sectional study. BMJ Open 2018 Jun 30;8(6):e020196. [doi: 10.1136/bmjopen-2017-020196] [Medline: 29961007]
6. Social networks and wellbeing. Australian Bureau of Statistics. 2016. URL: https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/4714.0~2014-15~Main%20Features~Social%20networks%20and%20wellbeing~4 [accessed 2024-08-05]
7. Fatima Y, Cleary A, King S, Solomon S, McDaid L, Hasan MM, et al. Cultural identity and social and emotional wellbeing in Aboriginal and Torres Strait Islander children. In: Baxter J, Lam J, Povey J, Lee R, Zubrick SR, editors. Family Dynamics over the Life Course Life Course Research and Social Policies: Springer; 2022, Vol. 15:57-70. [doi: 10.1007/978-3-031-12224-8_4]
8. Aboriginal and Torres Strait Islander people with a mental health condition. Australian Bureau of Statistics. 2016. URL: https://tinyurl.com/5xw32uzz [accessed 2024-08-05]
9. Determinants of health for First Nations people. Australian Institute of Health and Welfare. 2024. URL: https://www.aihw.gov.au/reports/australias-health/social-determinants-and-indigenous-health [accessed 2026-01-16]
10. Sabbioni D, Feehan S, Nicholls C, et al. Providing culturally informed mental health services to Aboriginal youth: the YouthLink model in Western Australia. Early Interv Psychiatry 2018 Oct;12(5):987-994. [doi: 10.1111/eip.12563] [Medline: 29573565]
11. Hepworth J, Askew D, Foley W, et al. How an urban Aboriginal and Torres Strait Islander primary health care service improved access to mental health care. Int J Equity Health 2015 Jun 6;14(1):51. [doi: 10.1186/s12939-015-0183-x] [Medline: 26048616]
12. Isaacs AN, Pyett P, Oakley‐Browne MA, Gruis H, Waples‐Crowe P. Barriers and facilitators to the utilization of adult mental health services by Australia's Indigenous people: seeking a way forward. Int J Mental Health Nurs 2010 Apr;19(2):75-82. [doi: 10.1111/j.1447-0349.2009.00647.x]
13. Markwick A, Ansari Z, Sullivan M, McNeil J. Social determinants and psychological distress among Aboriginal and Torres Strait islander adults in the Australian state of Victoria: a cross-sectional population based study. Soc Sci Med 2015 Mar;128:178-187. [doi: 10.1016/j.socscimed.2015.01.014] [Medline: 25618607]
14. Sherwood J. Colonisation - it's bad for your health: the context of Aboriginal health. Contemp Nurse 2013 Dec;46(1):28-40. [doi: 10.5172/conu.2013.46.1.28] [Medline: 24716759]
15. Thurber KA, Colonna E, Jones R, et al. Prevalence of everyday discrimination and relation with wellbeing among Aboriginal and Torres Strait Islander adults in Australia. Int J Environ Res Public Health 2021 Jun 18;18(12):6577. [doi: 10.3390/ijerph18126577] [Medline: 34207406]




XSL•FO
RenderX

16. Kirkbride JB, Anglin DM, Colman I, et al. The social determinants of mental health and disorder: evidence, prevention and recommendations. World Psychiatry 2024 Feb;23(1):58-90. [doi: 10.1002/wps.21160] [Medline: 38214615]
17. National Aboriginal and Torres Strait Islander health survey. Australian Bureau of Statistics. 2024. URL: https://www.abs.gov.au/statistics/people/aboriginal-and-torres-strait-islander-peoples/national-aboriginal-and-torres-strait-islander-health-survey/latest-release [accessed 2024-08-11]
18. Page IS, Ferrari AJ, Slade T, Anderson M, Santomauro D, Diminic S. Estimating the difference in prevalence of common mental disorder diagnoses for Aboriginal and Torres Strait Islander peoples compared to the general Australian population. Epidemiol Psychiatr Sci 2022 Jun 21;31:e44. [doi: 10.1017/S2045796022000233] [Medline: 35726614]
19. Coates D, Saleeba C, Howe D. Mental health attitudes and beliefs in a community sample on the central coast in Australia: barriers to help seeking. Community Ment Health J 2019 Apr;55(3):476-486. [doi: 10.1007/s10597-018-0270-8] [Medline: 29589218]
20. Hunter E. Disadvantage and discontent: a review of issues relevant to the mental health of rural and remote Indigenous Australians. Aust J Rural Health 2007 Apr;15(2):88-93. [doi: 10.1111/j.1440-1584.2007.00869.x] [Medline: 17441816]
21. Larson A. Rural health's demographic destiny. Rural Remote Health 2006;6(2):551. [Medline: 16623618]
22. Reifels L, Bassilios B, Nicholas A, et al. Improving access to primary mental healthcare for Indigenous Australians. Aust N Z J Psychiatry 2015 Feb;49(2):118-128. [doi: 10.1177/0004867414562046] [Medline: 25492971]
23. Garay J, Williamson A, Young C, et al. Aboriginal young people's experiences of accessibility in mental health services in two regions of New South Wales, Australia. Int J Environ Res Public Health 2023 Jan 18;20(3):1730. [doi: 10.3390/ijerph20031730] [Medline: 36767098]
24. Kavanagh BE, Corney KB, Beks H, Williams LJ, Quirk SE, Versace VL. A scoping review of the barriers and facilitators to accessing and utilising mental health services across regional, rural, and remote Australia. BMC Health Serv Res 2023 Oct 4;23(1):1060. [doi: 10.1186/s12913-023-10034-4] [Medline: 37794469]
25. McIntyre M, Ehrlich C, Barnett L, Kendall E. The experience of isolated practice for Indigenous mental health and addictions workers. Int J Ment Health 2022 Jan 2;51(1):77-94. [doi: 10.1080/00207411.2021.1984128]
26. Schultz R, Quinn S, Wilson B, Abbott T, Cairney S. Structural modelling of wellbeing for Indigenous Australians: importance of mental health. BMC Health Serv Res 2019 Jul 15;19(1):488. [doi: 10.1186/s12913-019-4302-z] [Medline: 31307436]
27. 3.10 Access to mental health services. Australian Institute of Health and Welfare. URL: https://www.indigenoushpf.gov.au/measures/3-10-access-to-mental-health-services [accessed 2026-01-16]
28. 3.14 Access to services compared with need. Australian Institute of Health and Welfare. URL: https://www.indigenoushpf.gov.au/measures/3-14-access-to-services-compared-with-need [accessed 2026-01-16]
29. Sveticic J, Milner A, De Leo D. Contacts with mental health services before suicide: a comparison of Indigenous with non-Indigenous Australians. Gen Hosp Psychiatry 2012;34(2):185-191. [doi: 10.1016/j.genhosppsych.2011.10.009] [Medline: 22154656]
30. Steering Committee for the Review of Government Service Provision. Report on government services 2020. : Productivity Commission; 2020 URL: https://www.pc.gov.au/ongoing/report-on-government-services/2020/data-downloads/ [accessed 2026-01-16]
31. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of technology-based strategies to promote engagement with digital interventions: a systematic review protocol. JMIR Res Protoc 2015 Apr 28;4(2):e47. [doi: 10.2196/resprot.3990] [Medline: 25921274]
32. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research... and three things we can do about them. Psychiatr Serv 2017 May 1;68(5):427-429. [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
33. Tassnim R, Tawfic Ja'far RA, Al-Tarawneh AMA. 2024 2nd International Conference on Cyber Resilience (ICCR): IEEE; 2024. [doi: 10.1109/ICCR61006.2024]
34. Mendes-Santos C, Nunes F, Weiderpass E, Santana R, Andersson G. Understanding mental health professionals' perspectives and practices regarding the implementation of digital mental health: qualitative study. JMIR Form Res 2022 Apr 12;6(4):e32558. [doi: 10.2196/32558] [Medline: 35412459]
35. Christensen H, Hickie IB. Using e-health applications to deliver new mental health services. Med J Aust 2010 Jun 7;192(S11):S53-S56. [doi: 10.5694/j.1326-5377.2010.tb03695.x] [Medline: 20528711]
36. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11. [doi: 10.1002/wps.20083] [Medline: 24497236]
37. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS ONE 2010 Oct 13;5(10):e13196. [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
38. Spence J, Titov N, Johnston L, Jones MP, Dear BF, Solley K. Internet-based trauma-focused cognitive behavioural therapy for PTSD with and without exposure components: a randomised controlled trial. J Affect Disord 2014 Jun;162:73-80. [doi: 10.1016/j.jad.2014.03.009] [Medline: 24767009]
39. Mental health services. Australian Institute of Health and Welfare. URL: https://www.aihw.gov.au/mental-health/overview/mental-health-services [accessed 2024-08-12]




XSL•FO
RenderX

40. Bernard A, de Ossorno Garcia S, Salhi L, John A, DelPozo-Banos M. Patterns of engagement in a digital mental health service during COVID-19: a cohort study for children and young people. Front Psychiatry 2023;14:1143272. [doi: 10.3389/fpsyt.2023.1143272] [Medline: 37575580]
41. Ftanou M, Machlin A, Mangelsdorf SN, Morgan A, Bassilios B. Australian digital mental health services: consumer perceptions of their usability and acceptability. J Technol Behav Sci 2024;10(2):320-335. [doi: 10.1007/s41347-024-00431-9]
42. Cheng L, Chan WK, Peng Y, Qin H. Towards data-driven tele-medicine intelligence: community-based mental healthcare paradigm shift for smart aging amid COVID-19 pandemic. Health Inf Sci Syst 2023 Dec;11(1):14. [doi: 10.1007/s13755-022-00198-4] [Medline: 36923686]
43. Dawson AZ, Walker RJ, Campbell JA, Davidson TM, Egede LE. Telehealth and indigenous populations around the world: a systematic review on current modalities for physical and mental health. Mhealth 2020;6:30. [doi: 10.21037/mhealth.2019.12.03] [Medline: 32632368]
44. Goetz CJ, Mushquash CJ, Maranzan KA. An integrative review of barriers and facilitators associated with mental health help seeking among Indigenous populations. Psychiatr Serv 2023 Mar 1;74(3):272-281. [doi: 10.1176/appi.ps.202100503] [Medline: 36065579]
45. Hicks LJ, Toombs E, Lund J, Kowatch KR, Hopkins C, Mushquash CJ. Expanding our understanding of digital mental health interventions for Indigenous youth: an updated systematic review. J Telemed Telecare 2025 Aug;31(7):919-960. [doi: 10.1177/1357633X241239715] [Medline: 38584397]
46. Moecke DP, Holyk T, Beckett M, et al. Scoping review of telehealth use by Indigenous populations from Australia, Canada, New Zealand, and the United States. J Telemed Telecare 2024 Oct;30(9):1398-1416. [doi: 10.1177/1357633X231158835] [Medline: 36911983]
47. Caffery LJ, Bradford NK, Wickramasinghe SI, Hayman N, Smith AC. Outcomes of using telehealth for the provision of healthcare to Aboriginal and Torres Strait Islander people: a systematic review. Aust N Z J Public Health 2017 Feb;41(1):48-53. [doi: 10.1111/1753-6405.12600] [Medline: 27868300]
48. Choukou MA, Maddahi A, Polyvyana A, Monnin C. Digital health technology for Indigenous older adults: a scoping review. Int J Med Inform 2021 Apr;148:104408. [doi: 10.1016/j.ijmedinf.2021.104408] [Medline: 33609927]
49. Weiner MF, Rossetti HC, Harrah K. Videoconference diagnosis and management of Choctaw Indian dementia patients. Alzheimers Dement 2011 Nov;7(6):562-566. [doi: 10.1016/j.jalz.2011.02.006] [Medline: 22055972]
50. Fraser S, Mackean T, Grant J, Hunter K, Towers K, Ivers R. Use of telehealth for health care of Indigenous peoples with chronic conditions: a systematic review. Rural Remote Health 2017;17(3):4205. [doi: 10.22605/RRH4205] [Medline: 28930638]
51. Katapally TR. Smart Indigenous youth: the smart platform policy solution for systems integration to address Indigenous youth mental health. JMIR Pediatr Parent 2020 Sep 25;3(2):e21155. [doi: 10.2196/21155] [Medline: 32975527]
52. Fleming TM, Dixon RS, Merry SN. 'It's mean!' The views of young people alienated from mainstream education on depression, help seeking and computerised therapy. Advances in Mental Health 2012 Feb;10(2):195-203. [doi: 10.5172/jamh.2011.10.2.195]
53. Walker R, Usher K, Jackson D, et al. Connection to... Addressing digital inequities in supporting the well-being of young Indigenous Australians in the wake of COVID-19. Int J Environ Res Public Health 2021 Feb 22;18(4):2141. [doi: 10.3390/ijerph18042141] [Medline: 33671737]
54. Clair MS, Murtagh DP, Kelly J, Cook J. Telehealth a game changer: closing the gap in remote Aboriginal communities. Med J Aust 2019 Apr;210 Suppl 6(S6):S36-S37. [doi: 10.5694/mja2.50036] [Medline: 30927470]
55. Gibson K, O'Donnell S, Coulson H, Kakepetum-Schultz T. Mental health professionals' perspectives of telemental health with remote and rural First Nations communities. J Telemed Telecare 2011;17(5):263-267. [doi: 10.1258/jtt.2011.101011] [Medline: 21824967]
56. Newman L, Bidargaddi N, Schrader G. Service providers' experiences of using a telehealth network 12 months after digitisation of a large Australian rural mental health service. Int J Med Inform 2016 Oct;94:8-20. [doi: 10.1016/j.ijmedinf.2016.05.006] [Medline: 27573307]
57. Caffery LJ, Bradford NK, Smith AC, Langbecker D. How telehealth facilitates the provision of culturally appropriate healthcare for Indigenous Australians. J Telemed Telecare 2018 Dec;24(10):676-682. [doi: 10.1177/1357633X18795764] [Medline: 30343658]
58. Povey J, Raphiphatthana B, Torok M, et al. An emerging framework for digital mental health design with Indigenous young people: a scoping review of the involvement of Indigenous young people in the design and evaluation of digital mental health interventions. Syst Rev 2023 Jul 1;12(1):108. [doi: 10.1186/s13643-023-02262-w] [Medline: 37393283]
59. Moffatt JJ, Eley DS. The reported benefits of telehealth for rural Australians. Aust Health Rev 2010 Aug;34(3):276-281. [doi: 10.1071/AH09794] [Medline: 20797357]
60. Indigenous data sovereignty. : Maiam nayri Wingara Indigenous Data Sovereignty Collective and the Australian Indigenous Governance Institute; 2018 URL: https://www.maiamnayriwingara.org/mnw-principles [accessed 2026-01-16]
61. Liu M, Schueller SM. Moving evidence-based mental health interventions into practice: implementation of digital mental health interventions. Curr Treat Options Psych 2023;10(4):333-345. [doi: 10.1007/s40501-023-00298-2]




XSL•FO
RenderX

62. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8(1):45. [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
63. Campbell M, McKenzie JE, Sowden A, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ 2020 Jan 16;368:l6890. [doi: 10.1136/bmj.l6890] [Medline: 31948937]
64. Harfield S, Pearson O, Morey K, et al. Assessing the quality of health research from an Indigenous perspective: the Aboriginal and Torres Strait Islander quality appraisal tool. BMC Med Res Methodol 2020 Apr 10;20(1):79. [doi: 10.1186/s12874-020-00959-3] [Medline: 32276606]
65. Hong QN, Fàbregues S, Bartlett G, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. EFI 2018;34(4):285-291. [doi: 10.3233/EFI-180221]
66. Bennett-Levy J, Singer J, DuBois S, Hyde K. Translating e-mental health into practice: what are the barriers and enablers to e-mental health implementation by Aboriginal and Torres Strait Islander health professionals? J Med Internet Res 2017 Jan 11;19(1):e1. [doi: 10.2196/jmir.6269] [Medline: 28077347]
67. Brown K, Toombs M, Nasir B, et al. How can mobile applications support suicide prevention gatekeepers in Australian Indigenous communities? Soc Sci Med 2020 Aug;258:113015. [doi: 10.1016/j.socscimed.2020.113015] [Medline: 32559573]
68. Dingwall KM, Povey J, Sweet M, et al. Feasibility and acceptability of the Aboriginal and Islander mental health initiative for youth app: nonrandomized pilot with First Nations young people. JMIR Hum Factors 2023 Jun 7;10:e40111. [doi: 10.2196/40111] [Medline: 37285184]
69. Dingwall KM, Puszka S, Sweet M, Nagel T. "Like drawing into sand": acceptability, feasibility, and appropriateness of a new e‐mental health resource for service providers working with Aboriginal and Torres Strait Islander people. Aust Psychol 2015 Feb 1;50(1):60-69. [doi: 10.1111/ap.12100]
70. Fletcher R, Hammond C, Faulkner D, et al. Stayin' on Track: the feasibility of developing Internet and mobile phone-based resources to support young Aboriginal fathers. Aust J Prim Health 2017 Sep;23(4):329-334. [doi: 10.1071/PY16151] [Medline: 28449728]
71. Kennedy M, Kumar R, Ryan NM, Bennett J, La Hera Fuentes G, Gould GS. Codeveloping a multibehavioural mobile phone app to enhance social and emotional well-being and reduce health risks among Aboriginal and Torres Strait Islander women during preconception and pregnancy: a three-phased mixed-methods study. BMJ Open 2021 Nov 24;11(11):e052545. [doi: 10.1136/bmjopen-2021-052545] [Medline: 34819285]
72. Lee KSK, Conigrave JH, Callinan S, et al. Asking about the last four drinking occasions on a tablet computer as a way to record alcohol consumption in Aboriginal and Torres Strait Islander Australians: a validation. Addict Sci Clin Pract 2019 May 1;14(1):15. [doi: 10.1186/s13722-019-0148-2] [Medline: 31039824]
73. Ma JS, Batterham PJ, Kõlves K, et al. Community expectations and anticipated outcomes for crisis support services—Lifeline Australia. Health Social Care Comm 2022 Sep;30(5):1775-1788. [doi: 10.1111/hsc.13557]
74. Perdacher E, Kavanagh D, Sheffield J, Healy K, Dale P, Heffernan E. Using the stay strong app for the well-being of Indigenous Australian prisoners: feasibility study. JMIR Form Res 2022 Apr 8;6(4):e32157. [doi: 10.2196/32157] [Medline: 35394444]
75. Povey J, Mills P, Dingwall KM, et al. Acceptability of mental health apps for Aboriginal and Torres Strait Islander Australians: a qualitative study. J Med Internet Res 2016 Mar 11;18(3):e65. [doi: 10.2196/jmir.5314] [Medline: 26969043]
76. Povey J, Sweet M, Nagel T, et al. Determining priorities in the Aboriginal and Islander mental health initiative for youth app second phase participatory design project: qualitative study and narrative literature review. JMIR Form Res 2022 Feb 18;6(2):e28342. [doi: 10.2196/28342] [Medline: 35179498]
77. Povey J, Sweet M, Nagel T, et al. Drafting the Aboriginal and Islander mental health initiative for youth (AIMhi-Y) app: results of a formative mixed methods study. Internet Interv 2020 Sep;21:100318. [doi: 10.1016/j.invent.2020.100318] [Medline: 32477884]
78. Puszka S, Dingwall KM, Sweet M, Nagel T. E-mental health innovations for Aboriginal and Torres Strait Islander Australians: a qualitative study of implementation needs in health services. JMIR Ment Health 2016 Sep 19;3(3):e43. [doi: 10.2196/mental.5837] [Medline: 27644259]
79. Raphiphatthana B, Sweet M, Puszka S, Dingwall K, Nagel T. Evaluation of a three-phase implementation program in enhancing e-mental health adoption within Indigenous primary healthcare organisations. BMC Health Serv Res 2020 Jun 23;20(1):576. [doi: 10.1186/s12913-020-05431-y] [Medline: 32576266]
80. Raphiphatthana B, Sweet M, Puszka S, Whitty M, Dingwall K, Nagel T. Evaluation of electronic mental health implementation in Northern Territory Services using the integrated "promoting action on research implementation in health services" framework: qualitative study. JMIR Ment Health 2020 May 26;7(5):e14835. [doi: 10.2196/14835] [Medline: 32452812]
81. Routledge K, Snijder M, Newton N, et al. Acceptability and feasibility of Strong & Deadly Futures, a culturally-inclusive alcohol and drug prevention program for Aboriginal and/or Torres Strait Islander secondary students. SSM - Mental Health 2022 Dec;2:100073. [doi: 10.1016/j.ssmmh.2022.100073]
82. Russell S, Quigley R, Strivens E, et al. Validation of the Kimberley Indigenous Cognitive Assessment short form (KICA-screen) for telehealth. J Telemed Telecare 2021 Jan;27(1):54-58. [doi: 10.1177/1357633X19860309] [Medline: 31284810]




XSL•FO
RenderX

83. Snijder M, Stapinski L, Ward J, et al. Strong and deadly futures: co-development of a web-based wellbeing and substance use prevention program for Aboriginal and Torres Strait Islander and non-Aboriginal adolescents. Int J Environ Res Public Health 2021 Feb 23;18(4):2176. [doi: 10.3390/ijerph18042176] [Medline: 33672190]
84. Tighe J, Shand F, McKay K, Mcalister TJ, Mackinnon A, Christensen H. Usage and acceptability of the iBobbly app: pilot trial for suicide prevention in Aboriginal and Torres Strait Islander youth. JMIR Ment Health 2020 Dec 1;7(12):e14296. [doi: 10.2196/14296] [Medline: 33258782]
85. Tighe J, Shand F, Ridani R, Mackinnon A, De La Mata N, Christensen H. Ibobbly mobile health intervention for suicide prevention in Australian Indigenous youth: a pilot randomised controlled trial. BMJ Open 2017 Jan 27;7(1):e013518. [doi: 10.1136/bmjopen-2016-013518] [Medline: 28132007]
86. Titov N, Schofield C, Staples L, Dear BF, Nielssen O. A comparison of Indigenous and non-Indigenous users of MindSpot: an Australian digital mental health service. Australas Psychiatry 2019 Aug;27(4):352-357. [doi: 10.1177/1039856218789784] [Medline: 30058351]
87. Veinovic M, Hill TY, Lavrencic L, et al. Telephone cognitive screening with older Aboriginal Australians: a preliminary study. Australas J Ageing 2023 Jun;42(2):311-316. [doi: 10.1111/ajag.13147] [Medline: 36330778]
88. Australian Institute of Health and Welfare. Digital Mental Health Resources for First Nations People: AIHW, Australian Government; 2023. URL: https://www.aihw.gov.au/getmedia/28f92ca3-ee7f-478a-828f-cfa2e00916e7/aihw-imh-20-digital-mental-health-resources-for-first-nations-people.pdf [accessed 2026-01-16]
89. Torous J, Bucci S, Bell IH, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335. [doi: 10.1002/wps.20883] [Medline: 34505369]
90. Li J, Brar A. The use and impact of digital technologies for and on the mental health and wellbeing of Indigenous people: a systematic review of empirical studies. Comput Human Behav 2022 Jan;126:106988. [doi: 10.1016/j.chb.2021.106988]
91. Hensel JM, Ellard K, Koltek M, Wilson G, Sareen J. Digital health solutions for Indigenous mental well-being. Curr Psychiatry Rep 2019 Jul 1;21(8):68. [doi: 10.1007/s11920-019-1056-6] [Medline: 31263971]
92. Balaratnasingam S, Chapman M, Chong D, et al. Advancing social and emotional well-being in Aboriginal and Torres Strait Islander Australians: clinicians' reflections. Australas Psychiatry 2019 Aug;27(4):348-351. [doi: 10.1177/1039856218789765] [Medline: 30234370]
93. Campbell ANC, Rieckmann T, Pavlicova M, et al. Culturally tailored digital therapeutic for substance use disorders with urban Indigenous people in the United States: a randomized controlled study. J Subst Use Addict Treat 2023 Dec;155:209159. [doi: 10.1016/j.josat.2023.209159] [Medline: 37690525]
94. Chen Q, Zhao Z, Bao J, Lin J, Li W, Zang Y. Digital empowerment in mental health: a meta-analysis of internet-based interventions for enhancing mental health literacy. Int J Clin Health Psychol 2024;24(3):100489. [doi: 10.1016/j.ijchp.2024.100489] [Medline: 39220119]
95. Sockalingam S, Arena A, Serhal E, Mohri L, Alloo J, Crawford A. Building provincial mental health capacity in primary care: an evaluation of a project ECHO mental health program. Acad Psychiatry 2018 Aug;42(4):451-457. [doi: 10.1007/s40596-017-0735-z] [Medline: 28593537]
96. Goodyear-Smith F, Martel R, Darragh M, Warren J, Thabrew H, Clark TC. Screening for risky behaviour and mental health in young people: the YouthCHAT programme. Public Health Rev 2017;38(1):20. [doi: 10.1186/s40985-017-0068-1] [Medline: 29450092]
97. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629. [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
98. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146. [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
99. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006 Jun 23;8(2):e10. [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
100. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. Lancet Digit Health 2020 Aug;2(8):e435-e440. [doi: 10.1016/S2589-7500(20)30142-4] [Medline: 32835201]
101. Black C, Frederico M, Bamblett M. "Healing through culture": Aboriginal young people's experiences of social and emotional wellbeing impacts of cultural strengthening programs. Child Abuse Negl 2024 Feb;148:106206. [doi: 10.1016/j.chiabu.2023.106206] [Medline: 37183121]
102. Alon N, Perret S, Cohen A, et al. Digital navigator training to increase access to mental health care in community-based organizations. PS (Wash DC) 2024 Jun 1;75(6):608-611. [doi: 10.1176/appi.ps.20230391]
103. Dunn P, Hazzard E. Technology approaches to digital health literacy. Int J Cardiol 2019 Oct 15;293:294-296. [doi: 10.1016/j.ijcard.2019.06.039] [Medline: 31350037]




XSL•FO
RenderX

104. St Clair M, Murtagh D. Barriers to telehealth uptake in rural, regional, remote Australia: what can be done to expand telehealth access in remote areas? Stud Health Technol Inform 2019 Aug 8;266:174-182. [doi: 10.3233/SHTI190791] [Medline: 31397320]
105. Featherstone D, Ormond-Parker L, Ganley L, et al. Mapping the digital gap: 2023 outcomes report. : ARC Centre of Excellence for Automated Decision-Making and Society; 2023 URL: https://apo.org.au/node/324397 [accessed 2026-01-16]
106. Mob Link. Institute for Urban Indigenous Health. URL: https://www.iuih.org.au/our-services/mob-link [accessed 2025-04-07]
107. Zhou X, Smith AC, Nelson C, et al. Disparities in access to and utilisation of Medicare subsidised telehealth mental health services by Aboriginal and Torres Strait Islander Australians in Southeast Queensland. PsyArXiv. Preprint posted online on May 19, 2025. [doi: 10.31234/osf.io/kjs42_v1]
108. Sinclair S, Waruk J, Morrisieau T, Cochrane C, Saulog A. Our data, our truths, our voice: reclaiming First Nations data sovereignty in Manitoba. IJPDS 2024;9(5). [doi: 10.23889/ijpds.v9i5.2731]

## Abbreviations

**AIMhi:** Aboriginal and Islander Mental Health Initiative
**DMH:** digital mental health
**KICA-screen:** Kimberly Indigenous Cognitive Examination-short form
**MMAT:** Mixed Methods Appraisal Tool
**MMSE:** Mini-Mental State Examination
**PICO:** population, intervention, comparison, outcome
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**TKICA-screen:** telehealth version of the Kimberly Indigenous Cognitive Examination-short form
**TMMSE:** telehealth version of the Mini-Mental State Examination

*Edited by N Cahill; submitted 09.Jul.2025; peer-reviewed by O Mbanefo, P Usuemerai; accepted 17.Dec.2025; published 27.Jan.2026.*

*Please cite as:*
*Zhai S, Goodman A, Smith AC, Diminic S, Zhou X*
*The Effectiveness, Facilitators, and Barriers of Digital Mental Health Services for First Nations People in Australia: Systematic Scoping Review*
*Interact J Med Res 2026;15:e80386*
*URL: https://www.i-jmr.org/2026/1/e80386*
*doi:10.2196/80386*

© Siyu Zhai, Andrew Goodman, Anthony C Smith, Sandra Diminic, Xiaoyun Zhou. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 27.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# The Usability of Continuous Monitoring Devices With Deterioration Alerting Systems in Noncritical Care Units: Scoping Review

Jo-Fan Pan[1], PhD; Dawn Dowding[1], PhD; David Wong[2], PhD; Ashley Scott[1], MSc; Qimeng Zhao[1], MSc

[1]University of Manchester, Jean McFarlane Building, Oxford Road, Manchester, United Kingdom

[2]University of Leeds, Leeds, United Kingdom

**Corresponding Author:**
Jo-Fan Pan, PhD
University of Manchester, Jean McFarlane Building, Oxford Road, Manchester, United Kingdom

## Abstract

**Background:** Delayed recognition of patient deterioration in a non–intensive care unit (ICU) setting contributes to serious adverse events. Continuous monitoring devices with alerting systems offer real-time data to support early detection, but their effectiveness depends on usability. While prior reviews focus on clinical outcomes, usability—defined by effectiveness, efficiency, and satisfaction—remains underexplored.

**Objective:** This study aims to scope the evidence related to the usability of continuous monitoring devices with deterioration alerting in noncritical adult care units.

**Methods:** A scoping review was conducted following the Joanna Briggs Institute methodology and reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines. A comprehensive search of MEDLINE, Embase, Emcare, Web of Science, and IEEE Xplore was performed for studies published up to November 2024. Title and abstract screening, full-text review, and data extraction were independently conducted by 2 reviewers. Studies were included if they (1) evaluated the usability—defined as effectiveness, efficiency, or satisfaction—of continuous monitoring devices; (2) focused on adult patients in non-ICU hospital settings; (3) used primary data; (4) were published in English; and (5) described how clinicians received alerts.

**Results:** The search identified 1284 papers, with 35 included. Most studies focused on postoperative patients in surgical wards, mainly from the United States and the Netherlands. Only 2 studies used mixed methods, and 10 reported clinician characteristics. While effectiveness (71%) and efficiency (74%) were widely studied, satisfaction (46%) and usability barriers (29%) received less attention.

**Conclusions:** Continuous monitoring devices with deterioration alerts may reduce rapid response team calls and ICU transfers, save time, and maintain acceptable alarm frequencies with high user satisfaction. However, usability challenges persist, including technical issues, alarm fatigue, patient discomfort, and limited training or workflow integration. This review mapped current use, usability, and barriers, categorized key usability factors for improvement, and identified the need for further research on clinician perspectives and broader health care settings to enhance generalizability.

***(Interact J Med Res 2026;15:e75713)*** doi:10.2196/75713

**KEYWORDS**

continuous monitoring; monitoring; wearable devices; early warning score; vital sign; alerts; alarms; deterioration

## Introduction

In 2022 - 2023, the United Kingdom recorded 16.4 million hospital admissions, compared to 33.7 million in the United States and 12.1 million in Australia [1-3]. Many of these patients may experience severe adverse events (SAEs), such as cardiac arrests or intensive care unit (ICU) admissions, during their hospital stay, with estimates ranging from 5% to 10% [4,5]. According to UK data, 23% of in-hospital SAEs resulting in sudden death are due to failures to recognize or respond to patient deterioration [6]. This issue is particularly concerning in non-ICU environments, where higher nurse-to-patient ratios require nurses to divide their attention among multiple patients, making timely identification of patient deterioration more challenging [7]. As van Galen et al [8] reported, 46% of 49 unplanned ICU transfers were linked to insufficient monitoring by nurses, highlighting the urgent need for enhanced monitoring strategies to support timely identification of at-risk patients in general wards.

The early warning score (EWS) is a widely used tool for early detection of patient deterioration in non-ICU settings [9]. It assigns weighted scores to vital signs (eg, respiratory rate, oxygen saturation, blood pressure, heart rate, temperature, and consciousness) to assess severity and guide interventions [9].



XSL•FO
RenderX

Studies have shown that the EWS can effectively identify patient deterioration, potentially reducing mortality rates in general wards [10,11]. The systematic review by Smith et al [11] of 21 retrospective studies in academic hospitals, primarily involving patients from medical and surgical wards, found that applying EWS was associated with a reduction in hospital mortality rates from nearly 50% to 40% and reduced cardiac arrest rates from 50% to 35%. However, studies also highlight that the effectiveness of the EWS can be compromised by inaccuracies in recorded data and its inability to capture immediate, real-time changes in vital signs [12,13]. In many general wards, nurses measure vital signs infrequently due to high patient-to-nurse ratios (especially compared with ICU settings) and intense workloads, which leave significant gaps during which critical signs of patient deterioration may be missed [12,13]. Incomplete and delayed data are associated with 19% of EWS inaccuracies in the United Kingdom and untimely responses in 75% of high-risk cases [14].

To help nurses improve the accuracy and timeliness of patient monitoring, real-time continuous monitoring systems with deterioration alerts have been proposed. Using wearable sensors or bedside monitors, these systems provide continuous data on key physiological parameters and generate alerts when out-of-range values are detected [15,16]. Multiple reviews have examined the effectiveness of these systems in non-ICU settings. Two reviews highlight positive outcomes: Downey et al [15] reviewed 24 studies involving over 40,000 patients and reported benefits such as reduced ICU transfers, shorter hospital stays, and significant cost savings. Similarly, Sun et al [16], in their meta-analysis of 14 studies, found that patients under continuous monitoring had a 39% lower risk of mortality compared to those with intermittent monitoring. However, other reviews have not found that continuous monitoring improves patient outcomes. Cardona-Morell et al [17] conducted a systematic review and meta-analysis of 22 clinical studies, involving sample sizes ranging from 16 to 64,661, and found no significant differences in the proportion of high-risk patients requiring urgent attention, preventing serious adverse events, cardiac arrests, or reducing ICU transfers. Similarly, the meta-analysis by Areia et al [18] reported potential reductions in ICU admissions and complications but found no differences compared to standard monitoring.

The differences in results across these reviews may be attributed to several factors. First, the design and functionality of continuous monitoring devices and their deterioration alerting systems vary, leading to differences in how alerts are generated and how clinicians interact with the devices. For example, Areia et al [18] focused solely on wearable devices, while Cardona-Morrell et al [17] included both wearable devices and bedside monitors. Second, the sample size and composition differed across studies. Downey et al [15] included 24 studies covering 40,274 patients and 59 ward staff across 9 countries, whereas Sun et al [16] analyzed 14 studies involving only 14,880 patients without including staff perspectives, resulting in a smaller and more limited participant base. Finally, these reviews primarily focused on device effectiveness, with limited attention to factors such as patient satisfaction and clinicians' adoption preferences, while all emphasized the need for future research on usability factors to enhance the devices' adoption and overall effectiveness [15-18]. Collectively, these findings highlight the limited focus on usability in current reviews, underscoring the need for further investigation into usability barriers and their impact on device usage.

Usability is a key factor in the adoption and successful use of health care technology, as it directly influences whether systems can achieve their intended goals, such as improving patient outcomes. Notably, these devices do not function in isolation. While they cannot make decisions, they provide essential information that shapes clinical judgment. Therefore, careful consideration of how these devices are integrated into clinical practice is critical. In this sense, the device's usability reflects both its inherent attributes (eg, battery life, sensor design, and interface clarity) and its application within the broader health care context, where clinicians interpret data, make treatment decisions, and adapt workflows accordingly. Poor usability, such as unclear interfaces, false alarms, or complex workflows, can lead to clinician disengagement, reducing trust and limiting the system's impact on clinical decisions and care processes [19].

According to the International Organization for Standardization (ISO), usability is defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [20]. ISO Standard 9241-11:2018 further breaks usability into 3 dimensions: effectiveness, which refers to the accuracy and completeness with which users achieve their goals; efficiency, the resources required to achieve these goals; and satisfaction, the comfort and acceptability of the system to its users [20]. As highlighted in the previous paragraph, studies on continuous monitoring and alerting systems in non-ICU settings often emphasize effectiveness while neglecting other critical factors, such as satisfaction and efficiency [15-18]. Therefore, this review aimed to map existing evidence on the usability of continuous monitoring devices with deterioration alerting systems.

## Methods

The review followed the Joanna Briggs Institute methodology for scoping reviews. A scoping review was chosen because it enables mapping of heterogeneous evidence across varied definitions of "usability," ensuring that all aspects of the concept are covered. The review also adhered to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines for transparent reporting.

The objectives of the review were framed using a Population, Concept, and Context framework as follows:

- Participants: the review includes studies involving clinicians working in non-ICU adult care settings, focusing on those who interact directly with continuous monitoring devices with deterioration alerting capabilities.
- Concept: the concept evaluated is the usability (effectiveness, efficiency, and satisfaction) of continuous monitoring devices that monitor vital signs and trigger



XSL•FO
RenderX

alerts. These devices must demonstrate how they improve clinical outcomes, enhance workflow efficiency, and increase user satisfaction.
- Context: the context includes hospital settings where these devices are applied in real-world non-ICU scenarios. Studies must describe how clinicians receive alerts and interact with these systems to ensure a comprehensive usability assessment.

No published protocol or registration was in place before this study.

## Search

The search was conducted in published electronic databases up to November 13, 2024, including MEDLINE, Embase, Evidence-Based Medicine Reviews, Web of Science, and IEEE Xplore databases. Gray literature was searched through Web of Science.

The search strategy, incorporating the Population, Concept, and Context framework and MeSH (Medical Subject Headings) terms, was refined using search keywords from previous reviews on continuous monitoring devices with deterioration alerting systems' capabilities [15-18,21].

A sample search strategy for MEDLINE was as follows:

("adult care unit*" OR "general ward*" OR hospitalization OR inpatient) AND (continuous* OR "real-time" OR remote* OR wearable* OR sensor* OR monitor*) AND ("early warning score*" OR "track and trigger" OR "deterioration alert*" OR "deterioration alarm*" OR "deterioration warn*")

The final search terms and database results are provided in Multimedia Appendix 1.

## Eligibility Criteria

The review included studies that evaluated the usability (of continuous monitoring devices with deterioration alerting, operated by clinicians in non-ICU hospital settings. Usability, as defined by ISO 9241-11:2018, consists of 3 key dimensions:

- Effectiveness: the accuracy and completeness with which users achieve their goals (eg, impact on ICU transfers, hospital stays, and patient outcomes).
- Efficiency: the resources required to achieve these goals (eg, alarm burden, false alert rates, and nursing workload).
- Satisfaction: the comfort and acceptability of the system to its users (eg, clinician and patient perceptions of comfort and acceptability).

The review focused on wearable sensors and bedside monitors that automatically track vital signs—including heart rate, blood pressure, respiratory rate, oxygen saturation, and temperature—at intervals of no more than 15 minutes. Studies were included if they described how clinicians received alerts, as this helps assess the impact of alerts and devices on clinical practice. Exclusion criteria omitted studies conducted in ICU settings; those in which clinicians—especially nurses, who are primarily responsible for patient monitoring—were blinded to alerts or not included in the research; studies involving participants under the age of 18 years; non-English publications; and review or protocol papers. The full criteria are provided in Textbox 1.

To ensure consistency in applying the inclusion and exclusion criteria during the study selection process, a dual-reviewer approach was used. The second reviewer independently evaluated 33% of studies at each review stage, including title/abstract screening and full-text review. If discrepancies exceeded 25%, the criteria were revised, and the screening process was repeated until the discrepancy rate fell below this threshold. Conflicts of less than 25% were resolved through consensus or consultation with a third reviewer. Additionally, reference lists of included studies were examined to identify further relevant studies meeting the inclusion criteria.

**Textbox 1.** Inclusion and exclusion criteria.

**Inclusion criteria:**

- Studies evaluating the usability (effectiveness, efficiency, and satisfaction) of continuous monitoring devices with deterioration alerting in hospital settings.
- Studies demonstrating the use of continuous monitoring devices with deterioration alerting in patient monitoring: alerts must be triggered by monitoring one or more vital signs such as heart rate, blood pressure, respiratory rate, oxygen saturation, or temperature.
- Studies that describe how clinicians receive alerts.

**Exclusion criteria:**

- Studies not published in English.
- Studies that did not apply continuous monitoring devices with deterioration alerting in non–intensive care unit (ICU) hospital settings or failed to clearly segregate data from non-ICU and ICU settings.
- Studies in which clinicians do not receive or acknowledge alerts: exclude studies where clinicians are kept unaware of alerts from monitoring devices (ie, clinicians were "blinded" to alerts, or the device did not generate alerts, unless clinicians had another way to receive the alerts, such as from ringing notifications or a central station).
- Studies stating that they involved participants younger than 18 years.
- Review or protocol papers: systematic reviews, meta-analyses, narrative reviews, and research protocols are excluded to focus solely on original research and first-hand data regarding device usability.



XSL•FO
RenderX

### Data Extraction

A structured data extraction form (Multimedia Appendix 2) was used to collect details on citations (author and year), study characteristics (participants, setting, design, methods, and country), device features (type, function, alerts, and application), and usability aspects (effectiveness, efficiency, satisfaction, and barriers).

First, a pilot data extraction phase involved both reviewers independently reviewing 2 included studies. Second, after independent extraction, both reviewers convened to discuss their findings. Discrepancies were resolved through discussion, and any unresolved issues were referred to a third reviewer. Third, following alignment on the pilot studies, the first reviewer extracted data from the remaining studies. The second reviewer then reviewed these extractions to verify consistency with the initial agreement and the standard definitions discussed. Remaining conflicts were resolved through consensus or consultation with a third reviewer.

### Synthesis

Studies were summarized and categorized based on key attributes, including publication year, country, study design, and research aim. First, these categorizations mapped the current research landscape and identified the distribution of studies across various contexts. For example, the percentage of studies conducted in different wards was reported to highlight where these devices are most frequently used. Second, the extracted data were categorized by common characteristics, such as device type and alert type. Finally, a narrative synthesis was conducted, thematically categorizing findings to align with the research question and objectives. Studies were grouped by their focus on usability (effectiveness, efficiency, and satisfaction) and barriers to usability, summarizing their findings, such as the accuracy of devices in detecting patient deterioration, categorized under effectiveness.

## Results

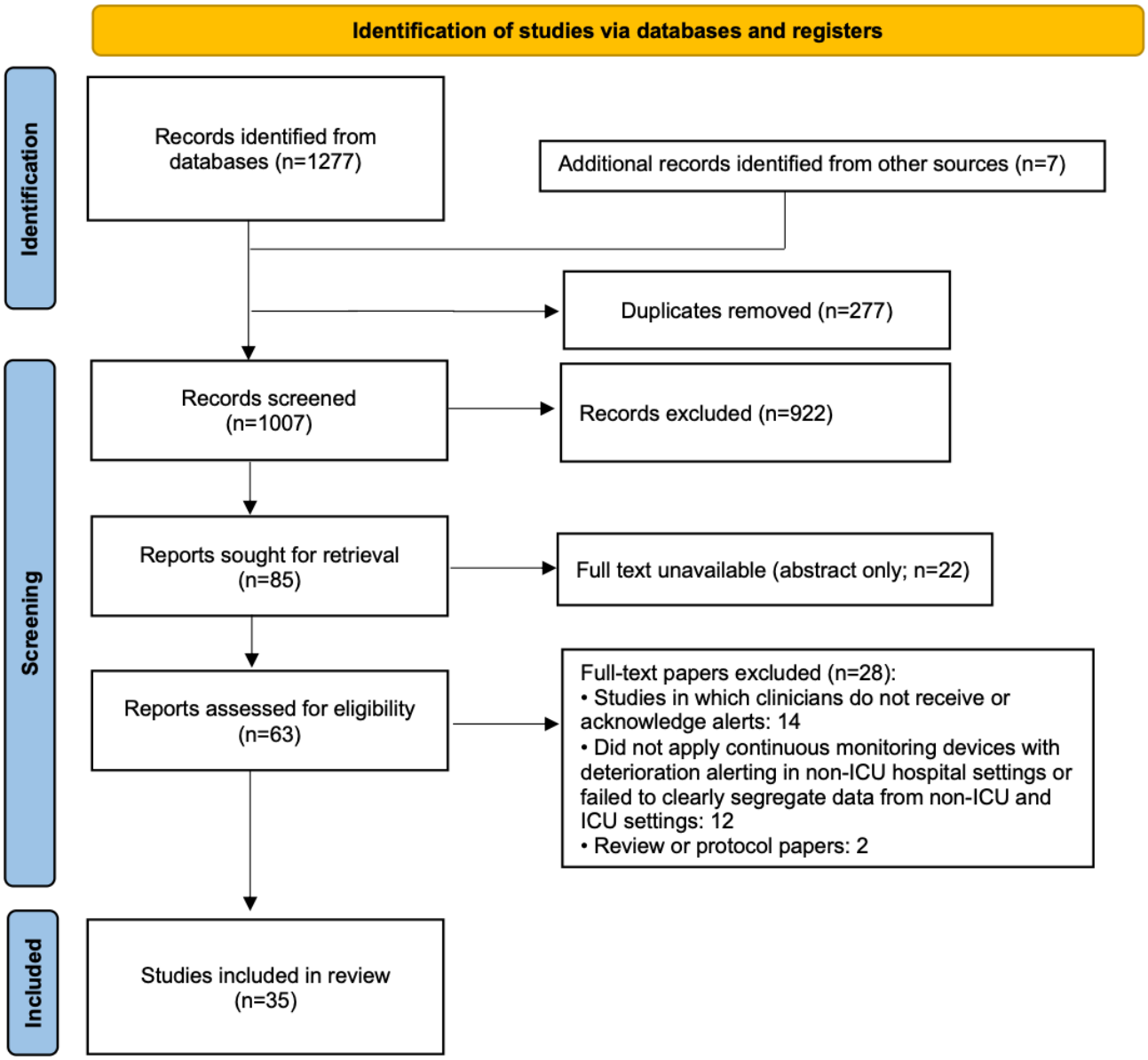
### Search Outcome

The search identified 1284 paper citations. After excluding duplicate records, 1007 records were deemed eligible for screening. A total of 63 studies were selected based on abstracts and underwent full-text review. After applying the inclusion and exclusion criteria, 35 studies were selected for this review. Figure 1 provides the study selection process in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (the PRISMA checklist is provided in Checklist 1).



XSL•FO
RenderX

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.




## Overview of Included Studies

An overview of the included studies, including study design, study devices, interventions, comparison groups, and outcomes measured, is summarized in Table 1. Overall, the 35 studies reported outcomes for a total of 65,029 patients and 323 clinicians in 8 countries. Multimedia Appendix 3 provides a summary of the included studies.



XSL•FO
RenderX

**Table .** Summary of the included studies.

| Author | Study aim type | Study design | Sample size | Alert mechanism type | Device type |
|---|---|---|---|---|---|
| Becking-Verhaar et al [22] | Implementation and feasibility of continuous monitoring systems | Cross-sectional survey | 58 nurses | Threshold alert | Wearable devices |
| Bellomo et al [23] | Impact on clinical outcomes and patient safety | Observational study | 18,305 patients (9617 before intervention and 8688 after intervention) | EWS[a] base alerts | Bedside monitors |
| Blankush et al [24] | Technological evaluation and alarm strategies | Observational study | 133 patients | EWS base alerts | Bedside monitors |
| Brown et al [25] | Impact on clinical outcomes and patient safety | Randomized controlled trial | Baseline cohort: 1535 (control) and 1433 (intervention); postimplementation cohort: 2361 (control) and 2314 (intervention) | Threshold alerts | Bedside monitors |
| Downey et al [26] | Nurses' and patients' perspectives and experiences | Randomized controlled trial | 226 patients randomized (140 to continuous monitoring and 86 to intermittent monitoring) | Threshold alerts | Wearable devices |
| Downey et al [27] | Comparison with episodic monitoring | Observational study | 12 patients | Threshold alerts | Wearable devices |
| Downey et al [28] | Implementation and feasibility of continuous monitoring systems | Randomized controlled trial | 136 patients | Threshold alerts | Wearable devices |
| Eddahchouri et al [29] | Comparison with episodic monitoring | Observational study | Baseline cohort: 2466 admissions and intervention cohort: 2303 admissions | Threshold alerts | Wearable devices |
| Gazarian [30] | Nurses' and patients' perspectives and experiences | Prospective, descriptive, observational study | 57 patients observed, 37 on continuous ECG[b] monitoring; 9 nurses | Threshold alerts | Wearable devices |
| Hravnak et al [31] | Technological evaluation and alarm strategies | Observational study | 629 patients (323 in phase I and 306 in phase III) | AI-based[c] alerts | Bedside monitors |
| Hravnak et al [32] | Technological evaluation and alarm strategies | Observational study | 326 patients | AI-based alerts | Bedside monitors |
| Joshi et al [33] | Technological evaluation and alarm strategies | Observational study | 50 patients | Threshold alerts | Wearable devices |
| Klumpner et al [34] | Technological evaluation and alarm strategies | Observational study | 64 monitored rooms | EWS base alerts | Bedside monitor + wearable devices |
| Kuznetsova et al [35] | Implementation and feasibility of continuous monitoring systems | Observational study | 35 (preimplementation: 13 and postimplementation: 22) clinicians | Threshold alerts | Bedside monitors |
| Leenen et al [36] | Implementation and feasibility of continuous monitoring systems | Observational study | 12 nurses | Threshold alerts | Wearable devices |
| Leenen et al [37] | Implementation and feasibility of continuous monitoring systems | Observational study | 30 patients and 23 nurses | Threshold alerts | Wearable devices |



XSL•FO
RenderX

| Author | Study aim type | Study design | Sample size | Alert mechanism type | Device type |
|---|---|---|---|---|---|
| McGrath et al [38] | Impact on clinical outcomes and patient safety | Observational study | 2 surgical units and 71 total beds | Threshold alerts | Wearable devices |
| McGrath et al [39] | Impact on clinical outcomes and patient safety | Observational study | Preimplementation: 4324 patient days and postimplementation: 4382 patient days | Threshold alerts | Wearable devices |
| McGrath et al [40] | Technological evaluation and alarm strategies | Observational study | General care units, including a 36-bed orthopedics unit and other surgical and medicine units; the exact number of participants was not specified, but the system covered more than 200 inpatient beds. | Threshold alerts | Wearable devices |
| Mestrom et al [41] | Impact on clinical outcomes and patient safety | Observational study | Control group: 320 patients and intervention group: 274 patients | EWS base alerts | Bedside monitor |
| Paul et al [42] | Impact on clinical outcomes and patient safety | Randomized controlled trial | Control group: 126 patients and intervention group: 124 patients | Threshold alerts | Wearable devices |
| Peelen et al [43] | Nurses' and patients' perspectives and experiences | Observational study | 1529 patients | AI-based alerts | Wearable devices |
| Pollack et al [44] | Implementation and feasibility of continuous monitoring systems | Observational study | 298 patients | Threshold alerts | Wearable devices |
| Posthuma et al [45] | Implementation and feasibility of continuous monitoring systems | Observational study | 742 patients (515 intermittent monitoring and 227 continuous monitoring) | Threshold alerts | Wearable devices |
| Sigvardt et al [46] | Nurses' and patients' perspectives and experiences | Observational study | 20 patients | Threshold alerts | Wearable devices |
| Stellpflug et al [47] | Impact on clinical outcomes and patient safety | Observational before-and-after study | 547 patients during the intervention period and 27 nurses | Threshold alerts | Wearable devices |
| Subbe et al [48] | Impact on clinical outcomes and patient safety | Observational before-and-after study | Control: 2139 patients and intervention: 2263 patients | Threshold alert | Bedside monitors + wearable devices |
| Taenzer et al [49] | Impact on clinical outcomes and patient safety | Observational before-and-after study | Preimplementation: 3118 discharges (intervention unit), 1260 (comparison unit 1), 2628 (comparison unit 2); postimplementation: 2841 discharges (intervention unit), 1162 (comparison unit 1), 2389 (comparison unit 2); 60 nurses | Threshold alert | Wearable devices |
| Un et al [50] | Implementation and feasibility of continuous monitoring systems | Observational study | 34 patients | AI-based alerts | Wearable devices |
| van Goor et al [51] | Implementation and feasibility of continuous monitoring systems | Observational before-and-after study | 209 patients (93 intermittent monitoring and 121 continuous monitoring) | Threshold alert | Wearable devices |



XSL•FO
RenderX

| Author | Study aim type | Study design | Sample size | Alert mechanism type | Device type |
|---|---|---|---|---|---|
| van Rossum et al [52] | Comparison with episodic monitoring | Observational retrospective study | 39 patients | Threshold alerts | Wearable devices |
| Verrillo et al [53] | Comparison with episodic monitoring | Observational study | Preintervention: 427 patients and intervention: 422 patients | Threshold alerts | Wearable devices |
| Watkins et al [54] | Impact on clinical outcomes and patient safety | Prospective observational study | 236 patients and 24 nurses | AI-based alerts | Wearable devices |
| Weenk et al [55] | Comparison with episodic monitoring | Randomized controlled trial | 60 patients | EWS-based alerts | Wearable devices |
| Weller et al [56] | Impact on clinical outcomes and patient safety | Prospective, observational study | 736 patients, 23 nurses, and 20 nursing assistants | Threshold alert | Wearable devices |

[a]EWS: early warning score.

[b]ECG: electrocardiogram.

[c]AI: artificial intelligence.

## Characteristics of Included Studies

The overview of the studies' characteristics is summarized in Table 2. The aim of the included studies was categorized into 5 main subjects: implementation and feasibility of continuous monitoring systems (n=9, 24%) [22,28,35-37,44,45,50,51], comparison with episodic monitoring (n=5, 14%) [27,29,52,53,55], impact on clinical outcomes and patient safety (n=11, 31%) [23,25,38,39,41,42,47-49,54,56], nurses' and patients' perspectives and experiences (n=4, 11%) [26,30,43,46], and technological evaluation and alarm strategies (n=6, 17%) [24,31-34,40]. The included studies were predominantly from 3 countries: the United States (n=16, 46%) [24,25,30-32,34,35,38-40,44,47,49,53,54,56], the Netherlands (n=10, 29%) [22,29,36,37,41,43,45,51,52,55], and the United Kingdom (n=5, 14%) [26-28,33,48]. The most common study design was observational (n=28, 80%) [23,24,27,29-41,43,44,46-54,56]. Additionally, 20 of the included studies (57%) were categorized as nonrandomized trials [23-25,29,31-33,38,39,41,43,44,47-52,54,56]. Mixed-method studies were the least frequently used, with 2 studies (6%) [22,40]. Seventy-four percent of the included studies had comparison groups (n=26), with before-and-after implementation comparisons (n=11, 31%) [23-25,31,32,35,39,40,49,51,56] and comparisons with intermittent monitoring (n=12, 34%) [26,28,29,34,38,41-43,45,46,48,53] being most commonly used. Sixty-three percent of studies used methods combining electronic health records (EHRs), observations, clinical trials, and experiments (n=22) [23-25,29-34,39,41-43,45,46,48-53,56]. The studies most commonly involved surgical-related wards (n=18, 51%) [22,24-27,29,30,35,36,38-40,42,43,45,49,52,54] and postoperative patients (n=17, 49%) [26-30,33,37-39,41-43,45,49,52-54].



XSL•FO
RenderX

**Table .** Overview of characteristics of included studies.

| Categories | Studies, n (%) | References |
|---|---|---|
| Study aim (n=35) | | |
| Implementation and feasibility of continuous | 9 (26) | [22,28,35-37,44,45,50,51] |
| Comparison with episodic monitoring | 5 (14) | [27,29,52,53,55] |
| Impact on clinical outcomes and patient safety | 11 (31) | [23,25,38,39,41,42,47-49,54,56] |
| Nurses' and patients' perspectives and experiences | 4 (11) | [26,30,43,46] |
| Technological evaluation and alarm strategies | 6 (17) | [24,31-34,40] |
| Country (n=35) | | |
| United States | 16 (46) | [24,25,30-32,34,35,38-40,44,47,49,53,54,56] |
| Netherlands | 9 (29) | [22,29,36,37,41,43,45,51,52,55] |
| United Kingdom | 5 (14) | [26-28,33,48] |
| United States, Europe, and Australia | 1 (3) | [23] |
| Others | 3 (9) | [42,46,50] |
| Study design (n=35) | | |
| Observational study (prospective and retrospective before-and-after) | 28 (71) | [23,24,27,29-41,43,44,46-54,56] |
| Randomized control trial | 6 (12) | [25,26,28,42,45,55] |
| Cross-sectional survey | 1 (3) | [22] |
| Comparison group (n=35) | | |
| Before-and-after implementation comparison | 11 (31) | [23-25,31,32,35,39,40,49,51,56] |
| Comparisons with intermittent monitoring | 12 (34) | [26,28,29,34,38,41-43,45,46,48,53] |
| Comparisons with baseline data | 1 (3) | [47] |
| Comparison between the monitor and alert method | 2 (6) | [52,55] |
| No comparison | 9 (26) | [22,27,30,33,36,37,44,50,54] |
| Clinical settings (n=35) | | |
| Surgical-related wards | 18 (51) | [22,24-27,29,30,35,36,38-40,42,43,45,49,52,54] |
| General wards | 7 (20) | [23,37,46-48,51,53] |
| Surgical and internal unit | 1 (3) | [55] |
| Others | 9 (26) | [28,31-34,41,44,50,56] |
| Patient type (n=35) | | |
| Postoperative | 17 (49) | [26-30,33,37-39,41-43,45,49,52-54] |
| General medical, trauma, and surgical patients | 1 (3) | [25] |
| Respiratory includes (COVID-19) | 4 (11) | [24,48,50,51] |
| Other | 6 (17) | [31,32,34,46,47,56] |
| Not specified | 7 (20) | [22,23,35,36,40,44,55] |
| Device type (n=35) | | |
| Wearable devices | 26 (74) | [22,26-30,33,36-40,42-47,49-56] |
| Bedside monitors | 5 (14) | [23-25,35,41] |
| Bedside monitors + wearable devices | 4 (11) | [31,32,34,48] |
| Alert path (n=35) | | |
| Alerts at central stations or system | 14 (40) | [22,23,29-32,35,38,41,43,44,47,50,54] |



XSL•FO
RenderX

| Categories | Studies, n (%) | References |
|---|---|---|
| Alert to the central station or system and clinicians' phones or mobile devices | 21 (60) | [24-28,33,34,36,37,39,40,42,45,46,48,49,51-53,55,56] |
| Alert mechanism type (n=35) | | |
| Threshold alerts | 25 (71) | [22,25-30,33,35-40,42,44-49,51-53,56] |
| Early warning score–based alerts | 5 (14) | [23,24,34,41,55] |
| Artificial intelligence–based alerts | 5 (14) | [31,32,43,50,54] |

## Characteristics of Clinicians

Ten (29%) of the included studies reported clinicians' characteristics involved in using continuous monitoring devices with deterioration alerting [22,30,35-37,47,49,53,54,56] (Multimedia Appendix 4). Seven studies reported incomplete characteristics [22,36,47,49,53,54,56], and 4 studies provided sample size and profession [47,49,54,56]. The available data indicated more female than male clinicians, with mean ages between 27 and 30 years. Most clinicians had 5-10 years of work experience, with nurses being the majority of users (Multimedia Appendix 5).

## Overview of Study Devices

The devices used can be classified into wearable devices and bedside monitors. Wearable devices, worn directly by the patient, monitor vital signs continuously while allowing patient mobility. The most commonly used wearable devices were the SensiumVital patch (Sensium Healthcare; n=8, 23%) [26-28,33,36,37,45,52] and ViSi Mobile (Sotera Wireless Inc; n=7, 20%) [22,43,51,53-56]. The SensiumVital patch, applied to the chest, monitors heart rate, respiratory rate, and body temperature, transmitting data every 2 minutes to a central station or mobile device, with visual alerts for deviations from preset vital signs [26]. ViSi Mobile monitors heart rate, blood pressure, respiratory rate, body temperature, and oxygen saturation from the upper arm, chest, and wrist, sending visual alerts to a central monitor and the nurse's Wi-Fi phone [56]. Bedside monitors are stationary devices placed near the patient's bed, offering continuous monitoring within the monitor's vicinity. The IntelliVue Guardian Solution (Philips), used in 4 studies, tracks heart rate, respiratory rate, blood pressure, body temperature, and oxygen saturation, providing visual alerts on central and bedside monitors [23,24,41,48]. It can also be used with wireless monitors on the chest, wrist, and upper arm, ensuring continuous monitoring and timely clinical actions [23,24,41,48].

The alerting mechanisms in the included studies were categorized into 3 groups: threshold alerts, EWS-based alerts, and artificial intelligence (AI)–based alerts. Threshold alerts (n=25, 71%) notify health care providers when a monitored vital sign exceeds predefined limits, typically applied to respiratory rate, oxygen saturation, heart rate, and blood pressure (with temperature less consistently included), although only a minority of studies reported explicit numerical cut-off values or which parameters most commonly triggered alarms [22,25-30,33,35-40,42,44-49,51-53,56]. EWS-based alerts (n=5, 14%) are generated using an aggregated score based on multiple vital signs to identify patients at risk of deterioration [23,24,34,41,55]. AI-based alerts (n=5, 14%) use algorithms to analyze vital signs and predict potential clinical deterioration [31,32,43,50,54]. In terms of delivery methods, 20 studies reported systems that sent alerts directly to clinicians' mobile devices, including phones or pagers [24-28,33,34,36,37,39,40,42,46,48,49,51-53,55,56], while also notifying the central nurse station. The remaining studies indicated that alerts were sent exclusively to the central station, highlighting variability in alert dissemination approaches.

## Effectiveness, Efficiency, Satisfaction, and Barriers to Usability

The included studies were analyzed for their reported effectiveness, efficiency, satisfaction, and barriers to usability.

## Effectiveness

Twenty-five studies (71% of all included studies) provided relevant data on effectiveness [23,25,26,28,29,31-33,36,37,40-45,47-50,52-56]. Eight of these studies (32%) reported on rapid response team (RRT) calls: 2 studies indicated an increase [23,48], 1 study reported no significant change [53], and 5 studies observed a decrease in RRT calls [29,40,47,49,56]. ICU transfer rates were evaluated in 13 (52%) studies, with 3 studies finding no significant change [25,42,45] and 10 studies reporting a decrease [28,29,31,32,40,47-49,53,56]. Mortality rates were assessed in 12 studies, with 5 studies noting an increase in survival rates [23,31,32,48,49] and 7 studies finding no significant change [26,28,29,41,53,56]. The length of hospital stay was examined in 12 studies; 2 studies did not specify results [37,50], 5 studies found no significant change [29,41,45,49,56], and 5 studies reported a decrease [23,25,26,28,47]. Readmission rates were evaluated in 3 studies, with 1 study [41] finding no significant change and 2 studies [26,28] observing a decrease. Serious adverse events were investigated in 6 studies, with 2 studies identifying adverse events [50,52] and 4 studies reporting a decrease [31,32,48,53]. Notably, 5 studies using AI-based alerting systems demonstrated potential benefits [31,32,43,50,54], including reduced mortality and ICU transfer rates [31,32] and improved identification and reduction of SAEs [31,32,50]. However, no significant tendency was observed across other measured outcomes. Of the 4 studies that combined bedside monitors with wearable devices [31,32,34,48], 3 reported strong effectiveness in reducing mortality, ICU transfers, and SAEs [31,32,48]. However, the effectiveness of bedside monitors alone remains unclear due to the limited number of studies.

Overall, more studies suggested that RRT calls and ICU transfer rates decrease after implementing continuous monitoring devices



XSL•FO
RenderX

with deterioration alerting systems. However, the effects on mortality, length of hospital stay, readmissions, and serious adverse events remain inconclusive.

### Efficiency

Twenty-six studies (74%) provided insights into efficiency issues, such as alarm frequency, false alert rate, workload impact, and time saving [22-24,30-40,42-44,46,47,49-52,54-56]. Alarm frequency was reported in 15 studies, with 4 studies indicating excessive alarms (over 4 alarms per patient per day) [24,30,36,56] and 11 studies reporting fewer than 5 alarms per patient per day [34,35,37-40,42,43,49,52,54]. The false alert rate was reported in 12 studies; 7 studies reported a high false alert rate (>40%) [24,30,36,37,43,50,52], while 5 studies reported a low false alert rate [31,34,44,49,56]. The impact on workload was investigated in 15 studies; 5 studies were unsure of the impact [22,39,47,51,56], 3 studies reported an increase in workload [24,36,54], and 7 studies observed a reduction in workload [32,34,35,38,40,46,50]. Time savings were evaluated in 8 studies, with 1 study reporting unclear results [55] and 7 studies reporting time savings [22,35,38,44,46,47,56]. Most current studies suggest that continuous monitoring devices save time, although the impact on workload is less clear. Alarm frequency generally remained below 5 alarms per patient per day, but false alarms were reported as a common and significant issue.

### Satisfaction

Satisfaction with the devices was reported in 16 (46%) studies. Comfort was reported in 8 studies, focusing only on patient perspectives. Among these, 88% (n=7) [26-28,37,42,44,47] reported that patients found the devices comfortable, while 1 study reported discomfort [22]. Acceptability was reported in 94% of the studies, with 8 studies reporting patient acceptability [22,26-28,37,39,42,49] and 9 studies reporting clinician acceptability [35,36,38,40,44,47,49,56]. High acceptability was found among both patients and clinicians, with 1 study not clearly stating clinician acceptability [53].

### Barriers to Usability

Usability barriers were formally reported in 10 studies (29% of all included studies) [22,26,27,30,35,36,39,44,47,49], while 5 studies mentioned them only in their discussion [24,28,40,42,43]. These barriers included technical issues, alarm management, patient comfort and experience, training and knowledge needs, and workflow integration challenges (Table 3).

**Table .** Summary of barriers to usability (n=10, 29% of all studies).

| Studies | Definition | Studies, n (%) |
|---|---|---|
| Technical issues [22,39,44] | Issues with device connectivity, battery life, sensor attachment, internet connectivity, and overall device design, along with technical difficulties and artifactual issues, such as poor lead adherence and interference, as well as the cumbersome nature of device use, including removal and reapplication for showers. | 3 (20) |
| Alarm management [26,30,49] | Frequent and high rates of initial false alarms leading to alarm fatigue and inconvenience, inconsistent practices in managing alarms, and excessive alarms from the system, often due to malfunctioning hardware or baseline tachycardia. | 3 (20) |
| Patient comfort and experience [26,27,36,39] | Discomfort and skin reactions from the patch, concerns about practicalities such as showering, trust issues with technology reliability, patient refusals due to discomfort, confusion, or personal reasons, and discontinued monitoring due to various factors, such as contact allergies or initiation of palliative care. | 4 (27) |
| Training and knowledge needs [22,26,35,36,39] | The need for patient education, ongoing training and coaching for health care providers, and challenges including alarm fatigue, accuracy and trust issues, and insufficient training for secondary users are impacting the patient experience. | 5 (33) |
| Workflow and integration issues [27,36,47,49] | Challenges with integrating the monitoring system into clinical workflows, managing alarm burdens, interpreting vital sign trends, ensuring seamless integration with hospital systems, and the potential reduction in face-to-face nursing contact. | 4 (27) |

Technical issues were identified in 3 studies and included problems with device connectivity, battery life, sensor attachment, internet connectivity, and overall device design, along with artefactual issues such as poor lead adherence (ie,



XSL•FO
RenderX

leads detaching frequently and affecting continuous monitoring), interference, and the cumbersome nature of device use, including removal and reapplication for showers [22,39,44]. Alarm management was a concern in 3 studies [26,30,49], with frequent and high rates of initial false alarms leading to alarm fatigue and inconvenience, inconsistent practices in managing alarms, and excessive alarms often resulting from malfunctioning hardware or baseline tachycardia. Unlike efficiency measures, which assess the total number of alarms and false alarm rates quantitatively, these studies highlighted usability-related challenges, such as how clinicians respond to alarms, their perceived reliability of alerts, and whether excessive alarms led to desensitization or delayed responses to actual critical alerts. Additionally, alarm issues were often linked to hardware malfunctions or patient baseline conditions (eg, persistent tachycardia triggering unnecessary alerts), which further complicated clinical workflows and increased clinician frustration. Patient comfort and experience were highlighted in 4 studies [26,27,36,39], discussing discomfort and skin reactions from the patch, practical concerns such as showering, trust issues with technology reliability, patient refusals due to discomfort, confusion, or personal reasons, and discontinued monitoring due to contact allergies or the initiation of palliative care. Training and knowledge needs were emphasized in 5 studies [22,26,35,36,39], underscoring the need for patient education, ongoing training, and coaching for health care providers, with challenges including alarm fatigue, accuracy and trust issues, and insufficient training for secondary users impacting the patient experience. Workflow and integration issues were reported in 4 studies [27,36,47,49], noting challenges with integrating the monitoring system into clinical workflows, managing alarm burdens, interpreting vital sign trends, ensuring seamless integration with hospital systems, and the potential reduction in face-to-face nursing contact. Finally, only 6 studies systematically surveyed usability barriers from clinicians' perspectives, collecting data through interviews, surveys, or by directly including clinicians as participants [22,30,35,36,47,49].

## Discussion

### Principal Findings

This scoping review aimed to map the evidence on the usability of continuous monitoring devices with deterioration alerting in non-ICU settings, according to the ISO standard. It focused on effectiveness, efficiency, satisfaction, and the barriers affecting usability. Current review findings suggest that while most research supports the effectiveness and efficiency of these devices, evidence regarding satisfaction and barriers to usability remains limited, with usability barriers receiving the least attention. Through the limited evidence, 5 key barriers to usability were identified in this review: (1) technical issues (eg, connectivity and battery limitations), (2) alarm management challenges (eg, false alarms and alarm fatigue), (3) patient comfort concerns (eg, skin irritation), (4) training gaps for clinicians, and (5) workflow integration difficulties. Considering the impact of barriers to usability on effectiveness, efficiency, and satisfaction, a critical gap in the literature is highlighted [19]. Future research should prioritize investigating usability barriers by examining patient and clinician experiences and developing interventions to overcome implementation challenges.

The findings of this review suggest that continuous monitoring devices with deterioration alerting are associated with reductions in RRT calls and ICU transfer rates in some studies, supporting previous reviews. Cardona-Morrell et al [17] and Sun et al [16] also reported significant reductions in cardiac arrest calls and rescue events associated with continuous monitoring. However, mortality outcomes remain inconclusive, likely due to variability in study designs, small sample sizes, and differences in patient populations or clinical settings. Similarly, Cardona-Morrell et al [17] and Areia et al [18] found no significant impact on mortality, while Sun et al [16] reported a 39% reduction in mortality risk, making it the only review with a significant finding. Interestingly, previous reviews all reported nonsignificant reductions in ICU transfers, which contrasts with the findings of this review [15-18]. These findings collectively suggest that continuous monitoring devices may improve patient outcomes; however, further research is needed to provide more evidence supporting their comprehensive benefits.

The efficiency of continuous monitoring devices with deterioration alerting is supported by evidence indicating time savings and manageable alarm frequencies. Most studies report fewer than 5 alarms per patient per day, which is generally acceptable to clinicians [56]. However, false alarms remain a significant concern, with some studies reporting rates exceeding 40% following device implementation. This issue has been highlighted in previous systematic reviews, which emphasize the need for improved alert accuracy, as false alarms are consistently identified as a usability barrier requiring further research and refinement [15,16,18]. Additionally, the impact of these devices on workload remains inconclusive, as findings vary across studies—some report a reduction in workload, others an increase, while some remain uncertain. Downey et al [15] noted that nurses who received proper training and felt confident using the technology experienced less workload strain, whereas a lack of familiarity led to disengagement and a perceived increase in workload. Given these inconsistencies, further research is needed to establish clearer trends regarding the workload impact of continuous monitoring devices.

Satisfaction with continuous monitoring devices is notably high, with both patients and clinicians reporting positive experiences. Most studies indicate that patients perceive these devices as comfortable and acceptable. While previous reviews have focused less on satisfaction metrics, Downey et al [15] provide valuable insights, noting that patients and clinicians recognize the clinical benefits of these devices and express willingness to adopt them due to enhanced patient safety. These findings align with the current review's conclusions regarding high satisfaction levels.

The barriers to usability associated with continuous monitoring devices with deterioration alerting systems in non-ICU settings are multifaceted, which prior reviews rarely address. Technical issues found in this review, such as unreliable Wi-Fi connectivity and sensors detaching from patients (triggering nonactionable alarms), align with findings from Leenen et al [21], whose review of 13 wearable devices highlighted design



XSL•FO
RenderX

limitations—including partial wiring in supposedly "wearable" systems that restrict patient mobility and generate clinically irrelevant alerts. In addition, patient discomfort and experience, manifested as skin irritation, mobility restrictions, and distrust in device reliability, have been widely recognized as a barrier in existing literature [15-18]. Furthermore, alarm management challenges, particularly false alarms, exacerbate clinician workload and desensitization, a concern echoed across previous review studies [15-18].

Notably, this review identifies training gaps and workflow integration challenges as critical yet underreported usability barriers that were not emphasized in earlier literature [15-18]. Inadequate education for both patients and clinicians—particularly regarding alarm interpretation, device operation, and troubleshooting—amplifies usability challenges and undermines system effectiveness. These findings align with Chaniaud et al [57], who suggest that even basic, widely adopted home devices (eg, blood pressure monitors and pulse oximeters) require robust user education to achieve their full potential. This highlights the universal importance of training programs tailored to device complexity and user expertise. Similarly, poorly integrated systems disrupt clinical workflows. For example, devices that lack seamless connectivity with existing EHR systems force clinicians to manually reconcile data, increasing staff burden and reducing time for direct patient care [49]. Collectively, these barriers highlight the necessity of holistic solutions that address not only technical performance but also human-centered design (eg, prioritizing patient comfort and clinician workflow efficiency) and system interoperability to maximize the potential of continuous monitoring.

One key aspect of the adoption and use of continuous monitoring systems is the critical role of clinicians who are responsible for implementing and managing these technologies. According to these findings, nurses make up the majority of front-line users, managing continuous monitoring, interpreting alerts, and overseeing these technologies in non-ICU environments. This highlights the need to prioritize nurses' needs in the design and integration of continuous monitoring with deterioration alert systems. However, a significant limitation in the literature is the disproportionate focus on patient experiences compared to clinician-related factors. This imbalance is problematic because clinicians' acceptance and effective use of technology are pivotal to integrating these systems successfully into clinical workflows. Few studies have examined usability barriers from the clinicians' perspective or provided detailed data on factors such as familiarity with devices, willingness to adopt new technologies, and demographic characteristics (eg, age, gender, and experience). In fact, clinicians' attributes can significantly influence usability, user behavior, and overall adoption of these systems [58]. Or et al [58] demonstrate that clinicians who trust and are willing to use technology are more likely to adopt EHR systems in their practice. These findings align with Carayon and Hoonakker [59], who argue that clinicians—as both implementers and end users—are as critical as system designers in determining health IT effectiveness. Their review underscores that neglecting clinician-specific barriers risks poor adoption and suboptimal outcomes. Addressing these gaps is essential for ensuring the long-term viability of continuous monitoring technologies and highlights the urgent need for studies that prioritize clinician-centered usability metrics.

Finally, this review adds several key points compared to previous reviews. First, AI-based alerting systems remain underused and are not yet well tested in these studies, indicating a need for further research to establish their effectiveness for future applications. A systematic review by Muralitharan et al [60] suggests that AI-based alerting methods perform better than threshold alerts and EWS-based alerts; however, this scoping review found limited practical application of AI-based alerting, underlining the necessity for additional clinical trials. Second, most studies on continuous monitoring devices with deterioration alerting are concentrated in the United States and are primarily conducted in surgical units with postoperative patients. This suggests that the application of these devices is currently limited to specific clinical settings, patient populations, and geographic regions. However, continuous monitoring systems have been highlighted as promising remote patient monitoring solutions that can save clinicians time by alleviating staffing burdens and improving patient safety—particularly in light of staffing shortages in both limited-, low-, and middle-income countries [61]. Third, implications for smaller community hospitals and lower-resource wards also warrant consideration. Most of the implementations identified in this review were in larger, well-resourced centers, whereas prior work has highlighted that higher nurse-to-patient ratios in general wards and human-related monitoring failures are important contributors to delayed recognition of deterioration [7,8]. In settings where 1 nurse cares for more patients and technical support is limited, the additional alarms and infrastructure required for continuous monitoring may therefore have different consequences for workload, alarm fatigue, and value for money; although earlier reviews have described potential cost savings in selected high-resource wards, economic outcomes were rarely reported in this review's included studies, underscoring the need for formal economic evaluations—particularly in community and resource-constrained hospitals—before large-scale implementation [15,61]. These gaps highlight the importance of expanding research to assess the generalizability of continuous monitoring technologies by exploring diverse clinical environments and rigorously evaluating emerging AI-based alerting systems.

## Limitations

This review is limited to studies published in English, potentially excluding relevant non-English publications and unpublished studies. The focus on usability (effectiveness, efficiency, and satisfaction) excludes studies that did not mention or examine these aspects. Only studies providing primary data were included, potentially omitting those reporting secondary data, such as systematic reviews and meta-analyses, or unpublished studies. Due to significant variations in methodology, objectives, and reported data among the included studies, a meta-analysis was not feasible. The search strategy excluded studies conducted in the ICU or nonhospital services, focusing deliberately on noncritical adult care units to investigate usability in these settings. This may have omitted relevant studies that include both ICU and non-ICU settings. Furthermore, as a scoping



XSL•FO
RenderX

review, this study aimed to include a broad range of studies without excluding any based on quality. Finally, usability was measured with heterogeneous, study-specific items, and no study used a validated instrument (eg, System Usability Scale), precluding cross-system benchmarking or pooled usability scores. Alert parameters and thresholds were inconsistently reported, limiting the comparability of alarm burden across devices and protocols.

## Conclusion

Continuous monitoring devices with deterioration alerting systems are increasingly recognized as valuable tools for preventing patient deterioration in non-ICU settings. However, their successful implementation hinges on a comprehensive understanding of usability (encompassing effectiveness, efficiency, and satisfaction) and the barriers influencing real-world adoption. This review indicates that these devices can reduce RRT calls and ICU transfers, save time, and maintain manageable alarm frequencies while achieving high user satisfaction. However, significant usability barriers remain, including technical issues, alarm management challenges, patient discomfort, and insufficient training and workflow integration. Moreover, most existing studies focus on effectiveness and efficiency, leaving satisfaction and broader usability factors understudied. Additionally, research has predominantly focused on patient perspectives, often neglecting clinician insights and has been limited to specific clinical contexts, patient populations, and geographic regions, which raises concerns about the generalizability of these findings. Future studies should prioritize usability factors and expand to the clinicians' usage perspective and diverse health care settings to ensure these technologies deliver equitable, scalable improvements in patient safety and optimize care delivery in non-ICU environments.

## Acknowledgments

This research was supported by a scholarship grant from the Government of Taiwan, awarded to JFP. The views expressed in this publication are those of the authors and do not necessarily reflect the official policy or position of their affiliated institutions.

## Funding

No external financial support or grants were received for this work.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search results up to November 2024.
[DOCX File, 26 KB - i-jmr_v15i1e75713_app1.docx ]

Multimedia Appendix 2
Data extraction sheet.
[DOCX File, 27 KB - i-jmr_v15i1e75713_app2.docx ]

Multimedia Appendix 3
Summary of the included studies.
[DOCX File, 78 KB - i-jmr_v15i1e75713_app3.docx ]

Multimedia Appendix 4
Characteristics of the included studies.
[DOCX File, 433 KB - i-jmr_v15i1e75713_app4.docx ]

Multimedia Appendix 5
Summary of the clinicians' characteristics in the included studies.
[DOCX File, 23 KB - i-jmr_v15i1e75713_app5.docx ]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 104 KB - i-jmr_v15i1e75713_app6.pdf ]

## References

1. Hospital admitted patient care activity, 2022-23. NHS England. 2023. URL: https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2022-23 [accessed 2026-02-04]



XSL•FO
RenderX

2. Total number of hospital admissions in the US from 1946 to 2022. Statista. URL: https://www.statista.com/statistics/459718/total-hospital-admission-number-in-the-us/?__sso_cookie_checker=failed [accessed 2026-01-30]
3. Admitted patient care. Australian Institute of Health and Welfare. URL: https://www.aihw.gov.au/reports-data/myhospitals/intersection/activity/apc [accessed 2026-01-30]
4. Massey D, Chaboyer W, Anderson V. What factors influence ward nurses' recognition of and response to patient deterioration? An integrative review of the literature. Nurs Open 2017 Jan;4(1):6-23. [doi: 10.1002/nop2.53] [Medline: 28078095]
5. Churpek MM, Yuen TC, Park SY, Gibbons R, Edelson DP. Using electronic health record data to develop and validate a prediction model for adverse outcomes in the wards*. Crit Care Med 2014 Apr;42(4):841-848. [doi: 10.1097/CCM.0000000000000038] [Medline: 24247472]
6. Donaldson LJ, Panesar SS, Darzi A. Patient-safety-related hospital deaths in England: thematic analysis of incidents reported to a national database, 2010-2012. PLoS Med 2014 Jun;11(6):e1001667. [doi: 10.1371/journal.pmed.1001667] [Medline: 24959751]
7. Bray K, Wren I, Baldwin A, et al. Standards for nurse staffing in critical care units determined by: The British Association of Critical Care Nurses, The Critical Care Networks National Nurse Leads, Royal College of Nursing Critical Care and In-flight Forum. Nurs Crit Care 2010;15(3):109-111. [doi: 10.1111/j.1478-5153.2010.00392.x] [Medline: 20500648]
8. van Galen LS, Struik PW, Driesen B, et al. Delayed recognition of deterioration of patients in general wards is mostly caused by human related monitoring failures: a root cause analysis of unplanned ICU admissions. PLoS ONE 2016;11(8):e0161393. [doi: 10.1371/journal.pone.0161393] [Medline: 27537689]
9. Gerry S, Bonnici T, Birks J, et al. Early warning scores for detecting deterioration in adult hospital patients: systematic review and critical appraisal of methodology. BMJ 2020 May 20;369:m1501. [doi: 10.1136/bmj.m1501] [Medline: 32434791]
10. Haegdorens F, Monsieurs KG, De Meester K, Van Bogaert P. An intervention including the national early warning score improves patient monitoring practice and reduces mortality: a cluster randomized controlled trial. J Adv Nurs 2019 Sep;75(9):1996-2005. [doi: 10.1111/jan.14034] [Medline: 31012124]
11. Smith MEB, Chiovaro JC, O'Neil M, et al. Early warning system scores for clinical deterioration in hospitalized patients: a systematic review. Ann Am Thorac Soc 2014 Nov;11(9):1454-1465. [doi: 10.1513/AnnalsATS.201403-102OC] [Medline: 25296111]
12. Williams TA, Tohira H, Finn J, Perkins GD, Ho KM. The ability of early warning scores (EWS) to detect critical illness in the prehospital setting: a systematic review. Resuscitation 2016 May;102:35-43. [doi: 10.1016/j.resuscitation.2016.02.011] [Medline: 26905389]
13. Zhang K, Zhang X, Ding W, et al. National Early Warning Score does not accurately predict mortality for patients with infection outside the intensive care unit: a systematic review and meta-analysis. Front Med (Lausanne) 2021;8:704358. [doi: 10.3389/fmed.2021.704358] [Medline: 34336903]
14. Kolic I, Crane S, McCartney S, Perkins Z, Taylor A. Factors affecting response to National Early Warning Score (NEWS). Resuscitation 2015 May;90:85-90. [doi: 10.1016/j.resuscitation.2015.02.009] [Medline: 25703784]
15. Downey CL, Chapman S, Randell R, Brown JM, Jayne DG. The impact of continuous versus intermittent vital signs monitoring in hospitals: a systematic review and narrative synthesis. Int J Nurs Stud 2018 Aug;84:19-27. [doi: 10.1016/j.ijnurstu.2018.04.013] [Medline: 29729558]
16. Sun L, Joshi M, Khan SN, Ashrafian H, Darzi A. Clinical impact of multi-parameter continuous non-invasive monitoring in hospital wards: a systematic review and meta-analysis. J R Soc Med 2020 Jun;113(6):217-224. [doi: 10.1177/0141076820925436] [Medline: 32521195]
17. Cardona-Morrell M, Prgomet M, Turner RM, Nicholson M, Hillman K. Effectiveness of continuous or intermittent vital signs monitoring in preventing adverse events on general wards: a systematic review and meta-analysis. Int J Clin Pract 2016 Oct;70(10):806-824. [doi: 10.1111/ijcp.12846]
18. Areia C, Biggs C, Santos M, et al. The impact of wearable continuous vital sign monitoring on deterioration detection and clinical outcomes in hospitalised patients: a systematic review and meta-analysis. Crit Care 2021 Dec;25(1). [doi: 10.1186/s13054-021-03766-4]
19. Coiera E. Assessing technology success and failure using information value chain theory. Stud Health Technol Inform 2019 Jul 30;263:35-48. [doi: 10.3233/SHTI190109] [Medline: 31411151]
20. Bevan N. International standards for HCI and usability. Int J Hum Comput Stud 2001 Oct;55(4):533-552. [doi: 10.1006/ijhc.2001.0483]
21. Leenen JPL, Leerentveld C, van Dijk JD, van Westreenen HL, Schoonhoven L, Patijn GA. Current evidence for continuous vital signs monitoring by wearable wireless devices in hospitalized adults: systematic review. J Med Internet Res 2020 Jun 17;22(6):e18636. [doi: 10.2196/18636] [Medline: 32469323]
22. Becking-Verhaar FL, Verweij RPH, de Vries M, Vermeulen H, van Goor H, Huisman-de Waal GJ. Continuous vital signs monitoring with a wireless device on a general ward: a survey to explore nurses' experiences in a post-implementation period. Int J Environ Res Public Health 2023 May 11;20(10):5794. [doi: 10.3390/ijerph20105794] [Medline: 37239523]




XSL•FO
RenderX

23. Bellomo R, Ackerman M, Bailey M, et al. A controlled trial of electronic automated advisory vital signs monitoring in general hospital wards. Crit Care Med 2012 Aug;40(8):2349-2361. [doi: 10.1097/CCM.0b013e318255d9a0] [Medline: 22809908]
24. Blankush JM, Freeman R, McIlvaine J, Tran T, Nassani S, Leitman IM. Implementation of a novel postoperative monitoring system using automated Modified Early Warning Scores (MEWS) incorporating end-tidal capnography. J Clin Monit Comput 2017 Oct;31(5):1081-1092. [doi: 10.1007/s10877-016-9943-4] [Medline: 27766526]
25. Brown H, Terrence J, Vasquez P, Bates DW, Zimlichman E. Continuous monitoring in an inpatient medical-surgical unit: a controlled clinical trial. Am J Med 2014 Mar;127(3):226-232. [doi: 10.1016/j.amjmed.2013.12.004]
26. Downey C, Randell R, Brown J, Jayne DG. Continuous versus intermittent vital signs monitoring using a wearable, wireless patch in patients admitted to surgical wards: pilot cluster randomized controlled trial. J Med Internet Res 2018 Dec 11;20(12):e10802. [doi: 10.2196/10802] [Medline: 30538086]
27. Downey CL, Brown JM, Jayne DG, Randell R. Patient attitudes towards remote continuous vital signs monitoring on general surgery wards: an interview study. Int J Med Inform 2018 Jun;114:52-56. [doi: 10.1016/j.ijmedinf.2018.03.014] [Medline: 29673603]
28. Downey CL, Croft J, Ainsworth G, et al. Trial of remote continuous versus intermittent NEWS monitoring after major surgery (TRaCINg): a feasibility randomised controlled trial. Pilot Feasibility Stud 2020 Nov 23;6(1):183. [doi: 10.1186/s40814-020-00709-8] [Medline: 33292669]
29. Eddahchouri Y, Peelen RV, Koeneman M, Touw HRW, van Goor H, Bredie SJH. Effect of continuous wireless vital sign monitoring on unplanned ICU admissions and rapid response team calls: a before-and-after study. Br J Anaesth 2022 May;128(5):857-863. [doi: 10.1016/j.bja.2022.01.036]
30. Gazarian PK. Nurses' response to frequency and types of electrocardiography alarms in a non-critical care setting: a descriptive study. Int J Nurs Stud 2014 Feb;51(2):190-197. [doi: 10.1016/j.ijnurstu.2013.05.014] [Medline: 23810495]
31. Hravnak M, Devita MA, Clontz A, Edwards L, Valenta C, Pinsky MR. Cardiorespiratory instability before and after implementing an integrated monitoring system. Crit Care Med 2011 Jan;39(1):65-72. [doi: 10.1097/CCM.0b013e3181fb7b1c] [Medline: 20935559]
32. Hravnak M, Edwards L, Clontz A, Valenta C, Devita MA, Pinsky MR. Defining the incidence of cardiorespiratory instability in patients in step-down units using an electronic integrated monitoring system. Arch Intern Med 2008 Jun 23;168(12):1300-1308. [doi: 10.1001/archinte.168.12.1300] [Medline: 18574087]
33. Joshi M, Ashrafian H, Arora S, et al. A pilot study to investigate real-time digital alerting from wearable sensors in surgical patients. Pilot Feasibility Stud 2022 Jul 6;8(1):140. [doi: 10.1186/s40814-022-01084-2] [Medline: 35794669]
34. Klumpner TT, Kountanis JA, Langen ES, Smith RD, Tremper KK. Use of a novel electronic maternal surveillance system to generate automated alerts on the labor and delivery unit. BMC Anesthesiol 2018 Jun 26;18(1):78. [doi: 10.1186/s12871-018-0540-6] [Medline: 29945569]
35. Kuznetsova M, Kim AY, Scully DA, et al. Implementation of a continuous patient monitoring system in the hospital setting: a qualitative study. Jt Comm J Qual Patient Saf 2024 Apr;50(4):235-246. [doi: 10.1016/j.jcjq.2023.10.017] [Medline: 38101994]
36. Leenen JPL, Dijkman EM, van Hout A, Kalkman CJ, Schoonhoven L, Patijn GA. Nurses' experiences with continuous vital sign monitoring on the general surgical ward: a qualitative study based on the Behaviour Change Wheel. BMC Nurs 2022 Mar 14;21(1):60. [doi: 10.1186/s12912-022-00837-x] [Medline: 35287678]
37. Leenen JPL, Dijkman EM, van Dijk JD, et al. Feasibility of continuous monitoring of vital signs in surgical patients on a general ward: an observational cohort study. BMJ Open 2021 Feb 17;11(2):e042735. [doi: 10.1136/bmjopen-2020-042735] [Medline: 33597138]
38. McGrath SP, Perreard IM, Garland MD, Converse KA, Mackenzie TA. Improving patient safety and clinician workflow in the general care setting with enhanced surveillance monitoring. IEEE J Biomed Health Inform 2019 Mar;23(2):857-866. [doi: 10.1109/JBHI.2018.2834863] [Medline: 29993903]
39. McGrath SP, Pyke J, Taenzer AH. Assessment of continuous acoustic respiratory rate monitoring as an addition to a pulse oximetry-based patient surveillance system. J Clin Monit Comput 2017 Jun;31(3):561-569. [doi: 10.1007/s10877-016-9884-y] [Medline: 27142098]
40. McGrath SP, Taenzer AH, Karon N, Blike G. Surveillance monitoring management for general care units: strategy, design, and implementation. Jt Comm J Qual Patient Saf 2016 Jul;42(7):293-302. [doi: 10.1016/s1553-7250(16)42040-4] [Medline: 27301832]
41. Mestrom E, De Bie A, Steeg MVD, et al. Implementation of an automated early warning scoring system in a surgical ward: practical use and effects on patient outcomes. PLoS ONE 2019;14(5):e0213402. [doi: 10.1371/journal.pone.0213402] [Medline: 31067229]
42. Paul JE, Chong MA, Buckley N, et al. Vital sign monitoring with continuous pulse oximetry and wireless clinical notification after surgery (the VIGILANCE pilot study)-a randomized controlled pilot trial. Pilot Feasibility Stud 2019;5(1):36. [doi: 10.1186/s40814-019-0415-8] [Medline: 30858986]




XSL•FO
RenderX

43. Peelen RV, Eddahchouri Y, Koeneman M, Melis R, van Goor H, Bredie SJH. Comparing continuous with periodic vital sign scoring for clinical deterioration using a patient data model. J Med Syst 2023 May 8;47(1):60. [doi: 10.1007/s10916-023-01954-z] [Medline: 37154986]
44. Pollack CV. Wireless cardiac event alert monitoring is feasible and effective in the emergency department and adjacent waiting areas. Crit Pathw Cardiol 2009 Mar;8(1):7-11. [doi: 10.1097/HPC.0b013e3181980f8b] [Medline: 19258832]
45. Posthuma LM, Breteler MJM, Lirk PB, Dijkum EJ, Visscher MJ, Breel JS, et al. Surveillance of high-risk early postsurgical patients for real-time detection of complications using wireless monitoring. Front Med (Lausanne) ;10. [doi: 10.3389/fmed.2023.1295499]
46. Sigvardt E, Grønbaek KK, Jepsen ML, et al. Workload associated with manual assessment of vital signs as compared with continuous wireless monitoring. Acta Anaesthesiol Scand 2024 Feb;68(2):274-279. [doi: 10.1111/aas.14333] [Medline: 37735843]
47. Stellpflug C, Pierson L, Roloff D, et al. Continuous physiological monitoring improves patient outcomes. Am J Nurs 2021 Apr 1;121(4):40-46. [doi: 10.1097/01.NAJ.0000742504.44428.c9] [Medline: 33755624]
48. Subbe CP, Duller B, Bellomo R. Effect of an automated notification system for deteriorating ward patients on clinical outcomes. Crit Care 2017 Dec;21(1). [doi: 10.1186/s13054-017-1635-z]
49. Taenzer AH, Pyke JB, McGrath SP, Blike GT. Impact of pulse oximetry surveillance on rescue events and intensive care unit transfers. Anesthesiology 2010 Feb;112(2):282-287. [doi: 10.1097/ALN.0b013e3181ca7a9b] [Medline: 20098128]
50. Un KC, Wong CK, Lau YM, et al. Observational study on wearable biosensors and machine learning-based remote monitoring of COVID-19 patients. Sci Rep 2021 Feb 23;11(1):4388. [doi: 10.1038/s41598-021-82771-7] [Medline: 33623096]
51. van Goor HMR, Eddahchouri Y, van Loon K, et al. Can continuous remote vital sign monitoring reduce the number of room visits to patients suspected of COVID-19: a quasi-experimental study. Int J Nurs Stud 2021 Mar;115:103868. [doi: 10.1016/j.ijnurstu.2020.103868] [Medline: 33465579]
52. van Rossum MC, Vlaskamp LB, Posthuma LM, et al. Adaptive threshold-based alarm strategies for continuous vital signs monitoring. J Clin Monit Comput 2022 Apr;36(2):407-417. [doi: 10.1007/s10877-021-00666-4] [Medline: 33575922]
53. Verrillo SC, Cvach M, Hudson KW, Winters BD. Using continuous vital sign monitoring to detect early deterioration in adult postoperative inpatients. J Nurs Care Qual 2019;34(2):107-113. [doi: 10.1097/NCQ.0000000000000350] [Medline: 30095509]
54. Watkins T, Whisman L, Booker P. Nursing assessment of continuous vital sign surveillance to improve patient safety on the medical/surgical unit. J Clin Nurs 2016 Jan;25(1-2):278-281. [doi: 10.1111/jocn.13102] [Medline: 26531215]
55. Weenk M, Koeneman M, van de Belt TH, Engelen LJLPG, van Goor H, Bredie SJH. Wireless and continuous monitoring of vital signs in patients at the general ward. Resuscitation 2019 Mar;136:47-53. [doi: 10.1016/j.resuscitation.2019.01.017] [Medline: 30685546]
56. Weller RS, Foard KL, Harwood TN. Evaluation of a wireless, portable, wearable multi-parameter vital signs monitor in hospitalized neurological and neurosurgical patients. J Clin Monit Comput 2018 Oct;32(5):945-951. [doi: 10.1007/s10877-017-0085-0] [Medline: 29214598]
57. Chaniaud N, Métayer N, Megalakaki O, Loup-Escande E. Effect of prior health knowledge on the usability of two home medical devices: usability study. JMIR Mhealth Uhealth 2020 Sep 21;8(9):e17983. [doi: 10.2196/17983] [Medline: 32955454]
58. Or C, Tong E, Tan J, Chan S. Exploring factors affecting voluntary adoption of electronic medical records among physicians and clinical assistants of small or solo private general practice clinics. J Med Syst 2018 May 29;42(7):121-112. [doi: 10.1007/s10916-018-0971-0] [Medline: 29845400]
59. Carayon P, Hoonakker P. Human factors and usability for health information technology: old and new challenges. Yearb Med Inform 2019 Aug;28(1):71-77. [doi: 10.1055/s-0039-1677907] [Medline: 31419818]
60. Muralitharan S, Nelson W, Di S, et al. Machine learning-based early warning systems for clinical deterioration: systematic scoping review. J Med Internet Res 2021 Feb 4;23(2):e25187. [doi: 10.2196/25187] [Medline: 33538696]
61. Holtz BE, Urban FA, Oesterle J, Blake R, Henry A. The promise of remote patient monitoring. Telemed J E Health 2024 Dec;30(12):2776-2781. [doi: 10.1089/tmj.2024.0521] [Medline: 39535888]


## Abbreviations

**AI:** artificial intelligence
**EHR:** electronic health record
**EWS:** early warning score
**ICU:** intensive care unit
**ISO:** International Organization for Standardization
**MeSH:** Medical Subject Headings
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews



XSL•FO
RenderX

**RRT:** rapid response team
**SAE:** severe adverse event

*Edited by M Balcarras; submitted 09.Apr.2025; peer-reviewed by E Ding, H El-Kechai; revised version received 10.Nov.2025; accepted 12.Jan.2026; published 10.Feb.2026.*

*Please cite as:*
*Pan JF, Dowding D, Wong D, Scott A, Zhao Q*
*The Usability of Continuous Monitoring Devices With Deterioration Alerting Systems in Noncritical Care Units: Scoping Review*
*Interact J Med Res 2026;15:e75713*
*URL: https://www.i-jmr.org/2026/1/e75713*
*doi:10.2196/75713*

© Jo-Fan Pan, Dawn Dowding, David Wong, Ashley Scott, Qimeng Zhao. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 10.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Review

# Machine and Deep Learning for Detection of Moderate-to-Vigorous Physical Activity From Accelerometer Data: Systematic Scoping Review

Yahua Zi[1], MSc; Sjors RB van de Ven[2], MSc; Eco JC de Geus[2], PhD; Peijie Chen[1], PhD

[1]School of Exercise and Health, Shanghai University of Sport, Shanghai, China

[2]Department of Biological Psychology, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

**Corresponding Author:**
Peijie Chen, PhD
School of Exercise and Health
Shanghai University of Sport
No.399 Changhai Rd
Shanghai, 200438
China
Phone: 86 21 5125 3145
Email: chenpeijie@sus.edu.cn

## Abstract

**Background:** Accurate monitoring of moderate-to-vigorous physical activity (MVPA) is critical for advancing public health research and personalized interventions. Traditional accelerometry methods, reliant on regression-derived intensity cut points, exhibit significant misclassification errors and poor generalizability to the free-living environment. Recent advancements in machine learning (ML) and deep learning (DL) offer promising alternatives for automated MVPA detection.

**Objective:** This scoping review synthesizes evidence on ML and DL techniques for MVPA estimation and prediction using accelerometer data, focusing on performance, algorithm bias, sensor configurations, and translational potential.

**Methods:** Following PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines, we conducted a systematic search across PubMed, IEEE Xplore, and Web of Science (February 1995-April 2025), supplemented by snowball citation tracking. Two independent reviewers screened titles, abstracts, and full texts against predefined inclusion criteria. Data from included studies were charted by one reviewer and verified by the other, extracting details on study characteristics, sensor configuration, ML and DL techniques, validation methods, and performance metrics. A narrative synthesis approach was used, guided by 6 research questions, to collate and summarize the findings. The synthesis process was rigorously reviewed by multiple authors to ensure consistency.

**Results:** Of 1938 screened studies, 40 met the inclusion criteria, with 4 studies added by follow-up manual searches. While traditional ML models (eg, random forest, support vector machine) achieved strong laboratory performance with $F_1$-score of 87.4%-100% and accuracy of 87.9%-100%, their real-world performance declined by 8.0%-13.3% in $F_1$-score and 6.6%-12.2% in accuracy, due to environment noise and device heterogeneity. DL architectures (eg, convolutional neural networks, transformers) achieved robust performance by leveraging raw signal dynamics with an $F_1$-score of 71.9%-79.8% and an accuracy of 87.9%-100% in free-living settings. Hybrid models (eg, convolutional neural networks and long short-term memory) demonstrated state-of-the-art performance ($F_1$-score 91.4%-98.4%, accuracy 97.7%-99.0%). Wrist-worn sensors dominated studies (30/40, 75%) and matched hip/thigh placements in lab settings (mean $F_1$-scores: 86.5%-88.6%), but multisensor configurations (wrist + hip) yielded the highest accuracy (89.7%). Key challenges included algorithmic bias reducing applicability in older adult populations, and impaired reproducibility, with only 42.5% (17/40) of studies sharing code and data. Emerging opportunities are noted for edge computing and hybrid models integrating contextual data.

**Conclusions:** ML and DL significantly enhance MVPA monitoring by automating feature extraction and improving adaptability to free-living variability. However, persistent gaps in generalizability, inconsistent validation protocols, and transparency deficits hinder translation. The findings support the need for future research to prioritize inclusive model training, standardized reporting frameworks, and open science practices to realize the equitable potential of artificial intelligence–driven physical activity assessment.



XSL•FO
RenderX

*(Interact J Med Res 2026;15:e76601)* doi:10.2196/76601

**KEYWORDS**

physical activity intensity; raw accelerometer data; wearable sensors; free-living validation; classification; estimation; sensor placement; machine learning; deep learning

## Introduction

Moderate-to-vigorous physical activity (MVPA) is defined as activities requiring specific metabolic equivalent of tasks (METs), such as 3 METs or 4 METs [1,2]. It is critical to preventive health, linked to reduced risks of cardiovascular disease [3,4], diabetes [5], and premature mortality [6]. Current guidelines, such as those from the World Health Organization, emphasize MVPA as a priority; for example, children and adolescents are advised to engage in MVPA with an average of 60 minutes per day across the week to improve health [7,8]. Additionally, accurate measurement of physical activity is critical for identifying the individual, environmental, and sociocultural determinants and evaluating the efficacy of intervention strategies. Accelerometer-based motion sensors, owing to their compact design, durability, and low cost, have emerged as the predominant tool for objective physical activity assessment in diverse populations [9-12].

Traditional accelerometry methods, though widely adopted, have historically been underused in research due to reliance on intensity-based cut points derived from linear regression models or receiver operating curves [13,14]. These approaches establish thresholds by predicting energy expenditure from accelerometer counts. However, proprietary count-based thresholds, such as Freedson's cut points [15], exhibit significant misclassification of activity intensity (eg, sedentary, light, moderate, vigorous intensity) of approximately 50% in adults [16] and 28%-45% in children and adolescents [17-20]. Such methods fail to account for biomechanical nuances (eg, energy expenditure differences between walking on flat terrain vs uphill terrain) or uncontrolled variables in free-living environments, such as nonexercise movements [21,22]. The proliferation of conflicting regression-derived cut points has further complicated cross-study comparisons [23]. While these thresholds remain standard for quantifying activity intensity, their inability to accurately predict intensity across diverse activities is increasingly acknowledged [22].

The advent of machine learning (ML) and deep learning (DL) has revolutionized intensity recognition by enabling feature extraction and classification from raw accelerometer signals [24,25]. Compared with traditional cut point methods, ML models (eg, random forests [RFs] and support vector machines [SVMs]) leverage time- and frequency-domain features from high-resolution triaxial data (eg, 30-100 Hz) to reduce energy expenditure errors by 25%-50% in school-age children [19,26]. More recently, DL architectures, such as a convolutional neural network (CNN) for local temporal pattern detection, a long short-term memory network (LSTM) for modeling activity sequences, Transformers for long-range dependency learning, and hybrid models (eg, convolutional neural network and long short-term memory [CNN-LSTM]), have further advanced the field. These models identify MVPA bouts by modeling temporal dependencies in continuous data streams [27].

Three distinct methodological approaches have emerged for MVPA detection: The first one is based on activity classification, which directly identifies MVPA from activity-specific movement patterns [28,29]. The second one is based on energy expenditure prediction from predefined MET thresholds (eg, ≥3 METs) [19,30]. The third one is based on an end-to-end DL architecture that automates hierarchical feature extraction from raw accelerometer signals to classify activity intensity directly or through energy expenditure estimation [31-33]. Hybrid models, such as CNN-LSTM, further enhance performance by integrating spatial feature extraction (via convolutions) with temporal modeling (via recurrent layers) to identify subtle biomechanical patterns (eg, stride variability during running) and contextual transitions between movements [34]. However, over 60% of models remain inaccessible due to unshared code or validation protocols, perpetuating a "new cut-point conundrum" that undermines cross-study comparability and clinical utility [35].

Other shortcomings further undermine progress in MVPA-specific research. First, lab-based findings fail to be generalized to real-world conditions. For example, RF achieves >90% accuracy in lab settings [36,37], but its free-living performance degrades dramatically to around 66% [38]. However, only 10% of studies validate models in the real world [39], limiting translational relevance. Second, disparities in validation protocols, such as settings (laboratory-controlled vs free-living environments), or device placement (hip vs wrist), complicate cross-study comparisons. For instance, models trained on hip-based ActiGraph data often underperform when applied to wrist-worn devices [40]. Third, ethical and reproducibility challenges, such as algorithmic bias against older adults or clinical cohorts, and limited code or data sharing, hinder the translation.

Several systematic reviews have explored the broader field of activity recognition using accelerometers and artificial intelligence (AI). However, a focused synthesis on AI-driven MVPA detection is lacking. Previous reviews have either focused on physical activity type detection in real-life conditions rather than intensity-specific thresholds [41], provided the general methodologies of human activity recognition using wearable sensors and ML without a systematic analysis of performance and bias in MVPA classification [42-44], examined the validation of accelerometer-based monitors using ML but not within the specific context of MVPA's lab-to-real-world gap [27], or highlighted the critical issue of accessibility and reproducibility of novel analytical models but not connected them to the development of equitable MVPA models [35]. Other reviews have focused on specific aspects, such as calibration techniques [45], sport-specific movements [46], or compared DL architectures like CNN and LSTM [47]. While 2 recent



XSL•FO
RenderX

reviews touch on predicting physical activity intensity from smartphones or smartwatches [48,49], they do not encompass the full spectrum of research-grade and wearable sensors, model architectures, and the critical synthesis of translational challenges presented here.

Therefore, this scoping review is the first to systematically scope and synthesize the literature exclusively on ML and DL techniques for MVPA intensity. We uniquely quantify the performance of MVPA detection methods as a function of the sensors used, sensor placement, target populations, feature extraction strategies, model architectures, lab-to-real-world settings, and look at possible algorithmic bias introduced by the restricted age and health status of the tested participants.

## Methods

### Overview

This scoping review follows the Arksey and O'Malley framework, which includes 5 key stages: identifying the research question (RQ), identifying relevant studies, selecting studies, charting the data, and collating, summarizing, and reporting the results. The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) was also consulted to ensure methodological rigor. EndNote X9 (Clarivate Analytics) was used for reference management, deduplication, and the screening process.

### Identify the Research Questions

This paper presents a scoping review that synthesizes advancements in ML- and DL-driven MVPA estimation and prediction from accelerometer data. The review aims to answer the following RQs:

- RQ1: What ML and DL techniques have been and are currently used for MVPA detection from accelerometer data?
- RQ2: How do accelerometer specifications (eg, sensor type, sampling rate), body placement (eg, wrist, hip, and thigh), and multisensor configurations influence model performance and generalizability?
- RQ3: What's the magnitude of the performance gap between laboratory-controlled and free-living environments, and what potential factors contribute to this disparity?
- RQ4: How do validation protocols vary across studies, and how do inconsistencies in these protocols limit cross-study comparability and clinical utility?
- RQ5: To what extent do current models exhibit biased results, preventing generalization to older adult or clinical populations?
- RQ6: What proportion of studies adhere to open science practices, and how do transparency gaps hinder reproducibility, scalability, and equitable deployment?

### Identify Relevant Studies

To ensure a comprehensive and focused literature search, we used a multi-step process. The search strategy was developed and refined in discussion with all the authors, who have specialized expertise in systematic review methodologies and database search strategies. Initially, we conducted a preliminary manual search to identify eligible studies and determine relevant databases and query terms. The search strategy included the following keywords and their combinations: "artificial intelligence" (eg, "machine learning" and "deep learning"), "accelerometer" (eg, "wearable device," "smartphone," "smartwatch," and "inertial measurement unit (IMU)"), and "moderate-to-vigorous physical activity" (eg, MVPA, "physical activity intensity," and "energy expenditure").

The comprehensive search was conducted across multiple databases, including PubMed, IEEE Xplore, and Web of Science. The search was conducted on April 4, 2025. To further enhance the comprehensiveness of our search, a manual citation search was conducted using reference lists of relevant studies to identify other potentially eligible studies. The detailed search strategies used to find relevant studies for this scoping review are described in Multimedia Appendix 1.

### Selection of Eligible Studies

We followed the steps outlined in the PRISMA (the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension guidelines) flow diagram to select eligible studies. The study selection process involved 2 independent reviewers (YZ and SRBVDV) who screened titles and abstracts, followed by full-text assessment. Any discrepancies were resolved through discussion. This process is illustrated in the PRISMA flow diagram. The inclusion criteria were as follows: (1) studies that applied ML or DL technique; (2) studies that used accelerometry, no matter any location, number of sensors, or any type of devices, such as accelerometers or smartphones or smartwatches; (3) studies that estimated of MVPA as the outcome; (4) studies that focused on human, any age group or any health status; and (5) peer-reviewed studies published in English.

The exclusion criteria were as follows: (1) studies that did not involve ML or DL techniques (2) studies that relied on multimodal sensor systems (eg, integrated heart rate monitors with accelerometers) or nonaccelerometric data (eg, video-based estimation); (3) studies that focused on nonhuman or nonphysical activity contexts, such as only differentiating sedentary behavior from nonsedentary activity; (4) studies that focused on general activity recognition or physical activity intensity classification without MVPA-specific analysis; (5) studies that focused on theoretical models without empirical validation; (6) studies that were not peer reviewed or reported in a non-English language; and (7) studies without full text available.

### Data Charting

Duplicates were identified and removed using the automated deduplication feature in EndNote X9, which was configured to define duplicates based on matching author, publication year, and title field. This automated process was followed by a manual check to ensure the thoroughness and accuracy of deduplication. Then, guided by the RQs, the following details were extracted from included studies: study characteristics (eg, author, publication year, population characteristics), sensor configuration (brand and model, placement, sampling rate), ML or DL techniques used to detect MVPA (features choosing



XSL•FO
RenderX

strategy, features selected, MVPA classification technique), ground truth validation of MVPA (via indirect calorimetry [IC] or direct observation [DO]), validating setting (lab or free-living conditions), classification performance metrics ($F_1$-score and accuracy), and code availability. In this review, only the $F_1$-score and accuracy for MVPA classification were extracted. In cases where these metrics were not explicitly reported in the primary studies, they were inferred from the provided confusion matrices using the standard functions. Accuracy, representing the proportion of total correct predictions, was calculated as:

The $F_1$-score, the harmonic mean of precision and recall, was calculated as

Where:

In which, TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

### Collating, Summarizing, and Reporting the Results

To answer the RQs, the results are organized into 6 sections: evolution of feature engineering and model architectures, task-specific insights, sensor performance, validation practices, algorithmic bias, and reproducibility crisis.

A narrative synthesis approach was used, guided by the predefined RQs. The extracted data were summarized quantitatively (using frequencies and percentages) and qualitatively (identifying key themes and trends). To ensure rigor and trustworthiness, the data charting and initial synthesis were performed by one author (YZ) and critically reviewed by the others (SRBVDV, EJCDG, and PC) for accuracy and consistency.

### Ethical Considerations

This scoping review synthesized findings from previously published research involving human participants. No new participants were recruited, and no new primary data were collected for this review. Consequently, separate ethical approval for this specific synthesis was not required. All original studies included in this review were expected to have obtained appropriate ethical approval from their relevant institutional review boards or ethics committees and informed consent from participants, consistent with ethical standards for human participation research involving sensor data. We noted that the majority of included studies explicitly reported ethical approval within their publications. For studies where an explicit ethics statement was not found in the publication, we acknowledge this limitation in reporting transparency. As this review analyzed results reported in published literature and did not involve direct access to or reanalysis of the raw accelerometer data from the original studies, specific data licenses or permissions beyond the published findings were not required.

## Results

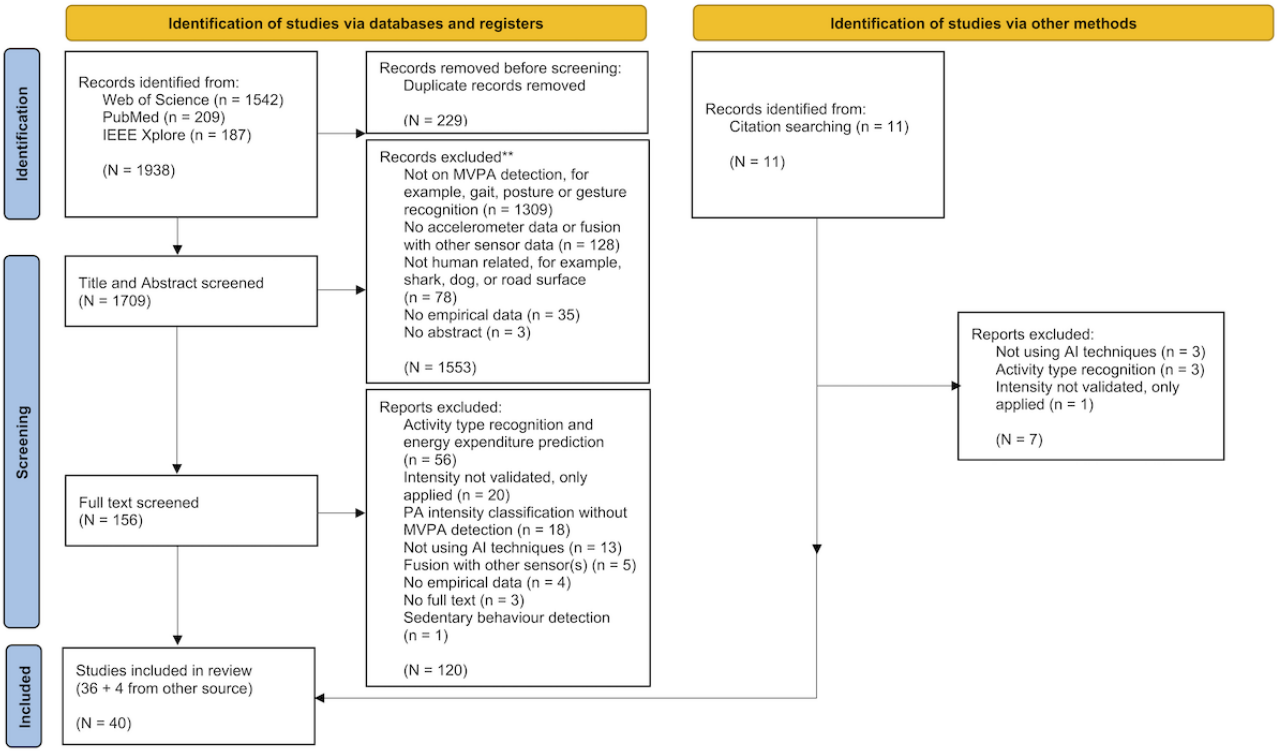
### Overview

A total of 1938 articles were identified from PubMed (n=209), IEEE Xplore (n=187), and Web of Science (n=1542). After removing 11.8% (229/1938) of duplicates, 88.2% (1709/1938) of the articles were included in the title and abstract screening phase. After this phase, 156 (8.1%) were screened for eligibility in the full-text screening phase. As a result, 36 articles met the inclusion criteria. In addition, 4 studies were included from manual searches. In total, 40 articles were included in this scoping review, shown in Figure 1.



XSL•FO
RenderX

**Figure 1.** PRISMA flow diagram of study selection. AI: artificial intelligence; MVPA: moderate-to-vigorous physical activity; PA: physical activity.




## Overview of Included Studies

Table 1 provides a summary of included studies, including the author's name and publication year, mean age of participants (SD), number of participants, sensor brand and specific model, number of sensors tested, the placement of sensors, MVPA classification techniques used, validation setting (lab or free-living), assessment of MVPA ground truth, and code availability. The information on the country where the test was conducted, sampling rate, feature choosing strategy, window length, feature selected, and $F_1$-score and accuracy of model detecting MVPA was listed in Multimedia Appendix 2 [19,28,29,31,33,34,36-38,50-81].



XSL•FO
RenderX

**Table 1.** The summary of included studies (N=40 studies, ranked by health condition and alphabetically by author names).

| Reference | Age (years), mean (SD) | Age (years), range | N | Sensor brand/ model | Number of sensors (placement) | MVPA[a] classification technique | Validation setting | Ground truth | Code availability |
|---|---|---|---|---|---|---|---|---|---|
| **Healthy condition** | | | | | | | | | |
| Ahmadi et al [50] | 4 (0.9) | —[b] | 31 | ActiGraph GT3X + | 2 (hip, wrist) | RF[c] and SVM[d] | Free-living | DO[e] | No |
| Ahmadi et al [51] | 4 (0.9) | — | 31 | ActiGraph GT3X + | 2 (hip, wrist) | RF | Free-living | DO | Yes |
| Ahmadi et al [52] | 13.9 (3) | — | 50 | ActiGraph GT3X + | 1 (hip) | RF | Lab | PRE[f] | Yes |
| Ahmadi et al [38] | 4 (0.9) | — | 31 | ActiGraph GT3X + | 2 (hip, wrist) | RF | Free-living | DO | Yes |
| Ahmadi et al [53] | 55.8 (12.4) | — | 102 | MEMS and ActiGraph GT9X | 2 (both wrists) | RF | Lab and free-living | DO | No |
| Ahmadi et al [53] | 55.8 (12.4) | — | 52 | MEMS and ActiGraph GT9X | 2 (both wrists) | RF | Lab and free-living | DO, IC[g] | No |
| Ahmadi et al [53] | — | 18-91 | 151 | Axivity AX3 | 1 (wrist) | RF | Free-living | DO | No |
| Andò et al [36] | NR[h] | — | NR | LSM9DS1 | 1 (chest) | k-NN[i] and RF | Lab | PRE | No |
| Bai et al [54] | 72.4 (7.1) | — | 247 | ActiGraph GT3X-BT | 1 (wrist) | XGBoost[j] | Lab | IC | No |
| Barua et al [34] | 29 (0) | 18-56 | 42 | Samsung Galaxy S7 | 3 (pocket, backpack, hand) | 1D-CNN-LSTM[k] | Lab | PRE | No |
| Chen et al [55] | 12.3 (1.0) | — | 18 | smartwatch “mumu” | 1 (wrist) | SVM | Free-living | IC | No |
| Chen et al [55] | 24.9 (2.6) | — | 24 | smartwatch “mumu” | 1 (wrist) | SVM | Free-living | IC | No |
| Davoudi et al [56] | 55.2 (17.8) | — | 40 | ActiGraph GT3X+ and Samsung smartwatch | 2 (wrist) | RF | Lab | IC | No |
| Doherty et al [57] | — | 18-91 | 153 | Axivity AX3 | 1 (wrist) | RF | Free-living | DO | No |
| Ellingson et al [58] | 23.9 (5.3) | — | 49 | ActiGraph GT3X+ and actiPAL | 2 (hip, thigh) | ANN[l] and DT[m] | Lab | IC and PRE | Yes |
| Ellingson et al [59] | 23.5 (4.6) | — | 51 | ActiGraph GT3X+ | 1 (hip) | RF | Lab | IC | On request |
| Farrahi et al [60] | 27.5 (11.2); 13.7 (3.1); 27.2 (3.3) | 20-30 | 22; 52; 9; 8 | Hookie AM20; Actigraph GT3X+; Colibri inertial measurement unit; Xsens MTx inertial measurement unit | 1 (hip); 2 (hip, wrist); 1 (wrist); 2 (wrists) | ANN | Lab | DO | No |
| Farrahi et al [31] | — | 18-91 | 151 | Axivity AX3 | 1 (wrist) | BiLSTM[n], RF, ANN, SVM, DT, and NB[o] | Free-living | DO | No |



XSL•FO
RenderX

| Reference | Age (years), mean (SD) | Age (years), range | N | Sensor brand/ model | Number of sensors (placement) | MVPA[a] classification technique | Validation setting | Ground truth | Code availability |
|---|---|---|---|---|---|---|---|---|---|
| Freedson et al [61] | 38 (12.4) | — | 277 | ActiGraph GT1M | 1 (hip) | ANN | Lab | IC | No |
| Hagenbuchner et al [29] | 4.8 (0.9) | — | 11 | ActiGraph GT3X + | 1 (hip) | ANN | Lab | PRE | No |
| Hibbing et al [62] | 9.4 (2.1) | — | 27 | ActiGraph GT3X-BT | 3 (hip, both wrists) | ANN and DT | Free-living | DO | On request |
| Hibbing et al [62] | 10.0 (2.2) | — | 54 | ActiGraph GT3X+ | 2 (hip, wrist) | ANN and DT | Lab | IC | On request |
| Li et al [63] | 4.0 (0.5) | — | 34 | ActiGraph GT3X-BT | 1 (wrist) | k-means[p] | Free-living | Hip cut points | No |
| Mardini et al [64] | 61.7 (17.7) | — | 253 | ActiGraph GT3X-BT | 1 (wrist) | DT, RF, XGBoost, and LASSO[q] | Lab | IC | No |
| Montoye et al [65] | 22.0 (4.2) | — | 40 | ActiGraph GT3X+ and GENEActiv | 4 (thigh, hip, both wrists) | ANN | Lab | PRE | No |
| Montoye et al [66] | 22.0 (4.2) | — | 41 | activPAL3 | 1 (thigh) | ANN | Lab | IC | No |
| Montoye et al [67] | 40.8 (19.2) | — | 48 | ActiGraph GT9X Link | 2 (hip, wrist) | RF | Lab and free-living | DO | Yes |
| Nawaratne et al [68] | 45.0 (11.0) | — | 119 | ActiGraph GT3X + | 1 (wrist) | CNN[r] | Free-living | Hip cut points | Yes |
| Nnamoko et al [69] | 69.3 (8.0) | — | 33 | GENEActiv and ActiGraph | 2 (wrist, hip) | Additive regression tree | Lab | IC | No |
| O'Driscoll et al [70] | 44.4 (14.1); 31.9 (10.2) | — | 89 | ActiGraph GT3-X; SenseWear Armband | 2 (wrist, upper arm) | RF, ANN, k-NN, SVM, and gradient boosting | Lab | IC | No |
| Pober et al [71] | 24.8 (4.2) | — | 6 | Actigraph MTI 7164 | 1 (hip) | QDA[s] and HMM[t] | Lab | PRE | No |
| Skjødt et al [72] | 80.2 (3.7) | — | 67 | ActiGraph GT3X +, GENEActiv, and Axivity AX3 | 6 (both hips, both wrists, thigh, lower back) | RF | Lab | IC | Yes |
| Staudenmayer et al [73] | 35 (0) | 21-69 | 48 | Actigraph model 7164 | 1 (wrist) | ANN | Lab | IC | No |
| Staudenmayer et al [74] | 24.1 (0) | 20-39 | 20 | ActiGraph GT3X+ | 1 (wrist) | RF and DT | Lab | IC | No |
| Trost et al [19] | 11 (2.7) | — | 100 | ActiGraph GT1M | 1 (hip) | ANN | Lab | IC | No |
| Trost et al [19] | 11 (2.7) | — | 100 | ActiGraph GT1M | 1 (hip) | ANN | Lab | IC | No |
| Trost et al [28] | 4.8 (0.9) | — | 11 | ActiGraph GT3X + | 2 (hip, wrist) | RF and SVM | Lab | DO, IC | On request |
| Tsanas [75] | — | 18-91 | 148 | Axivity AX3 | 1 (wrist) | RF and HMM | Free-living | DO | No |
| Walmsley et al [76] | — | 18-91 | 152 | Axivity AX3 | 1 (wrist) | RF and HMM | Free-living | DO | No |



XSL•FO
RenderX

| Reference | Age (years), mean (SD) | Age (years), range | N | Sensor brand/ model | Number of sensors (placement) | MVPA[a] classification technique | Validation setting | Ground truth | Code availability |
|---|---|---|---|---|---|---|---|---|---|
| Wang et al [33] | — | 18-91 | 151 | Axivity AX3 | 1 (wrist) | ViT-BiLSTM[u], CNN-BiLSTM[v], ViT[w], CNN, and BiLSTM | Free-living | DO | No |
| Wullems et al [77] | 73.5 (6.3) | — | 40 | GENEActiv | 2 (both thighs) | RF | Lab | IC | No |
| Wullems et al [78] | 70.0 (12.0) | — | 20 | GENEActiv | 1 (thigh) | RF | Lab | DO, IC | No |
| Zhou et al [79] | 5.0 (0.9) | — | 24 | Custom inertial measurement unit sensor | 1 (arm) | BiLSTM | Lab and free-living | IC | No |
| **Clinical conditions** | | | | | | | | | |
| Bianchim et al [37] | 12.0 (2.8) | — | $35^{CF}$; $28^{hy}$ | GENEActiv and ActiGraph | 5 (both wrists, waist, both wrists) | k-NN, RF, and XGBoost | Lab | IC | No |
| Cescon et al [80] | 44.9 (5.0) | — | $20^{T1Dz}$ | Empatica E4 wristband | 1 (wrist) | RF | Free-living | NR | No |

[a]MVPA: moderate-to-vigorous physical activity.
[b]Not applicable.
[c]RF: random forest.
[d]SVM: support vector machine.
[e]DO: direct observation.
[f]PRE: predefined activity schedule.
[g]IC: indirect calorimetry.
[h]NR: not reported.
[i]k-NN: k-nearest neighbor.
[j]XGBoost: extreme gradient boosting.
[k]1D-CNN-LSTM: one directional CNN-LSTM.
[l]ANN: artificial neural network.
[m]DT: decision tree.
[n]BiLSTM: bidirectional long short-term memory.
[o]NB: naive Bayes.
[p]k-means: k-means cluster analysis.
[q]LASSO: least absolute shrinkage and selection operator.
[r]CNN: convolutional neural network.
[s]QDA: quadratic discriminant analysis.
[t]HMM: hidden Markov model.
[u]ViT-BiLSTM: vision transformer bidirectional long short-term memory.
[v]CNN-BiLSTM: convolutional neural network and bidirectional long short-term memory.
[w]ViT: vision transformer.
[x]$35^{CF}$: 35 participants with cystic fibrosis.
[y]$28^{h}$: 28 healthy participants in the study.
[z]$20^{T1D}$: 20 participants with type 1 diabetes.

A total of 40 studies (2006-2025) met the inclusion criteria, with 62.5% (n=25) published between 2020 and 2025, reflecting the growing interest in AI-driven MVPA estimation.

Most studies (37/40 studies, 92.5%) targeted healthy populations, while only 5% (2/40 studies) addressed clinical cohorts, that is, cystic fibrosis [37] and type 1 diabetes [80], with one study (2.5%) did not specify the characteristics of participants [36].

Eleven studies (27.5%, 11/40) focused on children and adolescents [19,28,29,38,50-52,55,62,63,79], 40% (16/40) were



XSL•FO
RenderX

on adults (18-60 years old) [34,53,55,56,58,59,61,65-68,70,71,73,74], and 17.5% (7/40) were on old adults (60 years or older) [54,64,69,72,77,78], and 2 studies reported on the clinical conditions (ie, cystic fibrosis [37] and type 1 diabetes [80]). The remaining 7 studies (17.5%) tested their models using public datasets in adults, such as Capture-24 and Energy-24 (age range 18-91 years) [31,33,53,57,75,76] and a study with multiple datasets, including UOULU (University of Oulu), OSU (Oregon State University), the PAMAP2 Physical Activity Monitoring dataset (the UCI Machine Learning Repository), and the Daily and Sports Activities (the UCI Machine Learning Repository) [60]. Among these, Chen et al [55] covered the analyses both on children and adults; Andò et al [36], though not reporting participant ages, was contextually aligned with older adult research due to its emphasis on age-associated risks of physical inactivity among older adults and its heavy reliance on references related to older adults; Capture-24 and Energy-24 datasets [31,33,53,57,75,76] were grouped into adult, due to the age distributions: 72% of participants were younger than 53 years, with only 22.5% aged 53 years or older [82]; Farrahi et al [60], which included 4 datasets with an average participant age of about 19 years, was classified under adults.

ActiGraph (30/40, 75%) and GENEActiv (6/40, 15%) were the most common sensors using acceleration sensors to identify MVPA, with limited use of consumer wearables, for example, other brands of accelerometers (eg, Axivity AX3, activPAL; 9/40, 22.5%), inertial measurement units (5/40, 12.5%), smartwatches (2/40, 5%), smartphones (1/40, 2.5%), wristbands (1/40, 2.5%), and armbands (1/40, 2.5%).

Lab-controlled validations predominated (30/45 analyses, 66.7%; some studies had multiple analyses); 33.3% of analyses (15/45) were conducted in free-living conditions, while 4 analyses combined lab and free-living validations.

## Evolution of Feature Engineering and Model Architectures

A total of 45 analyses from 40 studies used a range of ML and DL techniques for MVPA detection.

### *Methodological Evolution*

The shift from feature-driven ML to end-to-end DL reflects a broader trend toward scalability and generalizability. While traditional ML models excel in interpretability and low computational cost, their dependence on handcrafted features renders them brittle in free-living contexts. In contrast, DL architectures, though data-hungry and opaque, inherently adapt to signal variability through hierarchical abstraction, a critical advantage for real-world deployment [31].

The progression from manual feature engineering to automated DL underscores a paradigm shift toward scalable, context-aware MVPA monitoring. Table 2 synthesizes this evolution, contrasting supervised, unsupervised, and hybrid paradigms. Supervised DL models, particularly those using transfer learning, now dominate, with all studies adopting pretrained CNN or bidirectional long short-term memory (BiLSTM) to mitigate data scarcity [31,33,34,68,79]. Unsupervised approaches, such as the self-organizing map and k-means cluster analysis, remain nascent but offer potential for leveraging unlabeled free-living data [83,84].



XSL•FO
RenderX

**Table 2.** Task-specific performance comparison.

| Task type and methods/ model | Key features | Performance metrics (number of studies) | References |
|---|---|---|---|
| **Classification (n=28)** | | | |
| RF[a] (n=13) | Handcrafted features: time/frequency features (eg, mean, SD, percentiles, lag-1 autocorrelation), ensemble of decision trees | • Lab (n=7): F1-score 91.9%, accuracy 94.0%<br>• Free-living (n=4): F1-score 81.0%, accuracy 87.4%<br>• Lab and free-living (n=2): F1-score 88.1%, accuracy 93.8% | [28,36-38,50-53,64,67,72,74,80] |
| ANN[b] (n=7) | Handcrafted features: time/frequency features (eg, spectral entropy, signal power), multilayer perceptron | • Lab (n=7): F1-score 88.0%, accuracy 93.1%<br>• Free-living (n=1): F1-score 75.4%, accuracy 82.1% | [19,29,58,60,62,65,73] |
| SVM[c] (n=4) | Kernal-based classification on RBF[d], advanced cross-correlation metrics (xy, xz, yz) | • Lab (n=1): F1-score 75.4%, accuracy 88.4%<br>• Free-living (n=3): accuracy 86.5% | [28,50,55] |
| DT[e] (n=4) | Tree-based splits, integrate with ANN outcomes | • Lab (n=3): F1-score 86.6%, accuracy 87.8%<br>• Free-living (n=1): F1-score 75.4%, accuracy 82.1% | [58,62,64,74] |
| Gradient boosting (n=3) | Gradient boosting framework, handling missing data | • Lab (n=2): F1-score 91.6% | [37,54,64] |
| HMM[f] (n=3) | Temporal sequence modeling, Viterbi smoothing | • Lab (n=1): F1-score 99.8%<br>• Free-living (n=2): F1-score 73.5%, accuracy 94.0% | [71,75,76] |
| QDA[g] (n=1) | Quadratic decision boundaries, probabilistic classification | • Lab (n=1): F1-score 100%, accuracy 99.9% | [71] |
| LASSO[h] (n=1) | L1 regularization, sparse solutions | • Lab (n=1): F1-score 83.6% | [64] |
| CNN[i] (n=1) | Automated feature extraction via convolutional filters on raw signals | • Free-living (n=1): F1-score 73.4%, accuracy 96.8% | [68] |
| **Estimation (n=10)** | | | |
| RF (n=6) | Regression trees, bootstrapped subsets of ActiGraph data | • Lab (n=5): F1-score 83.5%, accuracy 86.1%<br>• Free-living (n=1): F1-score 80.0%, accuracy 91.4% | [56,57,59,70,77] |
| ANN (n=2) | Nonlinear activation functions, raw signal processing | • Lab (n=2): F1-score 91.1%, accuracy 85.7% | [61,66] |
| SVM (n=1) | Kernal-based regression. | • Lab (n=1): F1-score 90.7%, accuracy 88.7% | [70] |
| k-NN[j] (n=2) | Instance-based learning, Euclidean distance metrics | • Lab (n=2): F1-score 96.4%, accuracy 95.8% | [37,70] |
| XGBoost[k] (n=1) | Gradient boosting framework, handling missing data | • Lab (n=1): F1-score 100%, accuracy 100% | [37] |
| Gradient boosting (n=1) | Iterative error correction, additive regression trees | • Lab (n=1): F1-score 93.2%, accuracy 92.1% | [70] |
| **Deep learning (n=5)** | | | |



XSL•FO
RenderX

| Task type and methods/ model | Key features | Performance metrics (number of studies) | References |
|---|---|---|---|
| Bi-LSTM[l] (n=3) | Bidirectional temporal modeling, raw signal processing | • Free-living (n=2): F1-score 73.6%, accuracy 93.6%<br>• Lab and free-living: F1-score 53.3%, accuracy 53.7% | [31,33,79] |
| CNN (n=2) | Automated feature extraction via convolutional filters on raw signals | • Free-living (n=2): F1-score 71.9%, accuracy 94.4% | [33,68] |
| ViT[m] (n=1) | Self-attention mechanisms for long-range dependencies | • Free-living (n=1): F1-score 79.8%, accuracy 95.0% | [33] |
| CNN-LSTM[n] or CNN-BiLSTM[o] (n=2) | Hybrid architecture, integrate spatial and temporal learning | • Lab (n=1): F1-score 82.1%<br>• Free-living (n=1): F1-score 91.4%, accuracy 97.7% | [33,34] |
| ViT-BiLSTM[p] (n=1) | Vision Transformer + BiLSTM, gravity-based acceleration analysis. | • Free-living (n=1): F1-score 98.4%, accuracy 99.0% | [33] |

[a]RF: random forest.

[b]ANN: artificial neural network.

[c]SVM: support vector machine.

[d]RBF: radial basis function.

[e]DT: decision tree.

[f]HMM: hidden Markov model.

[g]QDA: quadratic discriminant analysis.

[h]LASSO: least absolute shrinkage and selection operator.

[i]CNN: convolutional neural network.

[j]k-NN: k-nearest neighbor.

[k]XGBoost: extreme gradient boosting.

[l]BiLSTM: bidirectional long short-term memory.

[m]ViT: vision transformer.

[n]CNN-LSTM: convolutional neural network and bidirectional long short-term memory.

[o]CNN-BiLSTM: convolutional neural network and bidirectional long short-term memory.

[p]ViT-BiLSTM: vision transformer bidirectional long short-term memory.

### *Traditional Machine Learning*

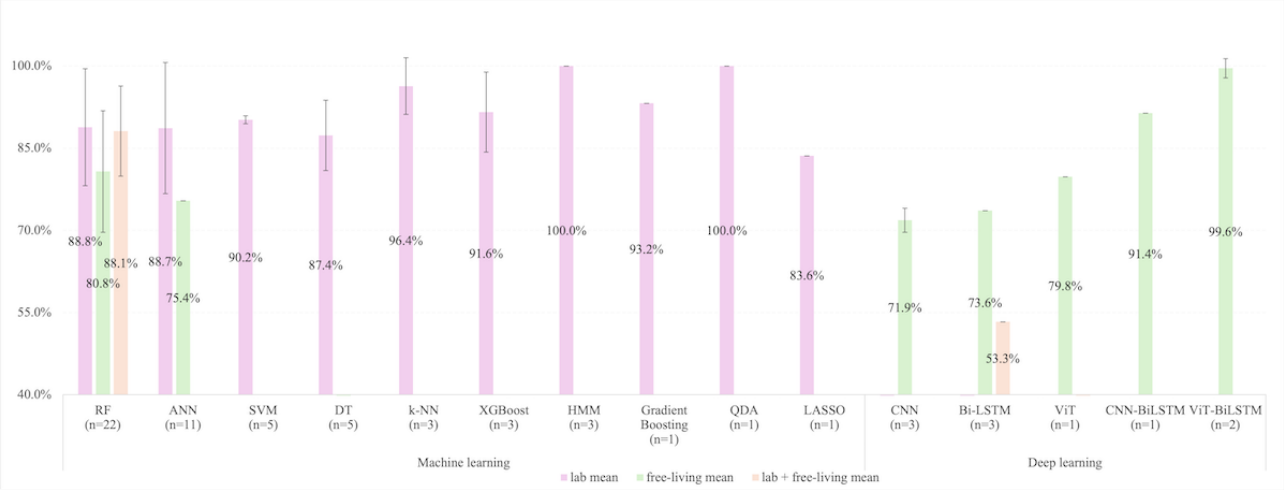
Traditional ML techniques have dominated accelerometer-based MVPA detection since 2006 [71], relying on handcrafted features derived from time- and frequency-domain analyses.

Among these, RF emerged as the most prevalent algorithm (22/40 studies, 55%) [28,36-38,50-53,56,57,59,64,67,70,72,74-78,80], achieving mean $F_1$-scores of 86.6% and mean accuracy of 88.6%. RF's ensemble structure, which aggregates predictions from multiple decision trees (DTs) (usually 100-1000), mitigates overfitting and enhances robustness to noise, a critical advantage in heterogeneous accelerometer datasets [67,76]. Artificial neural network (ANN) followed (11/40 studies, 27.5%), with mean $F_1$-scores of 87.4% and mean accuracy of 89.5% [19,29,31,58,60-62,65,66,70,73]. ANN used a multi-layer perceptron with input, hidden (3-25 nodes), and output layers, nonlinear activation functions model complex feature interactions. Another often-used method was the SVM (5/40 studies, 12.5%), with an $F_1$-score of 90.2% and accuracy of 86.5% [28,31,50,55,70]. SVM maps features to a high-dimensional space and constructs optimal hyperplanes using kernel functions (eg, radial basis function). The $F_1$-scores of models using DT (5/40 studies, 12.5%; $F_1$-score 87.4%) [31,58,62,64,74] underperformed k-nearest neighbor (3/40 studies, 7.5%; $F_1$-score 96.4%) [36,37,70], extreme gradient boosting (3/40 studies, 7.5%; $F_1$-score 91.6%) [54,64,80], hidden Markov model (3/40 studies, 7.5%; $F_1$-score 100%) [71,75,76], Gradient Boosting (1/40 studies, 2.5%; $F_1$-score 93.2%) [70], and quadratic discriminant analysis (1/40 studies, 2.5%; $F_1$-score 100%) [71]. Least absolute shrinkage and selection operator achieved the lowest $F_1$-score (83.6%) in detecting MVPA among all the ML models [64]. The details are illustrated in Figure 2.



XSL•FO
RenderX

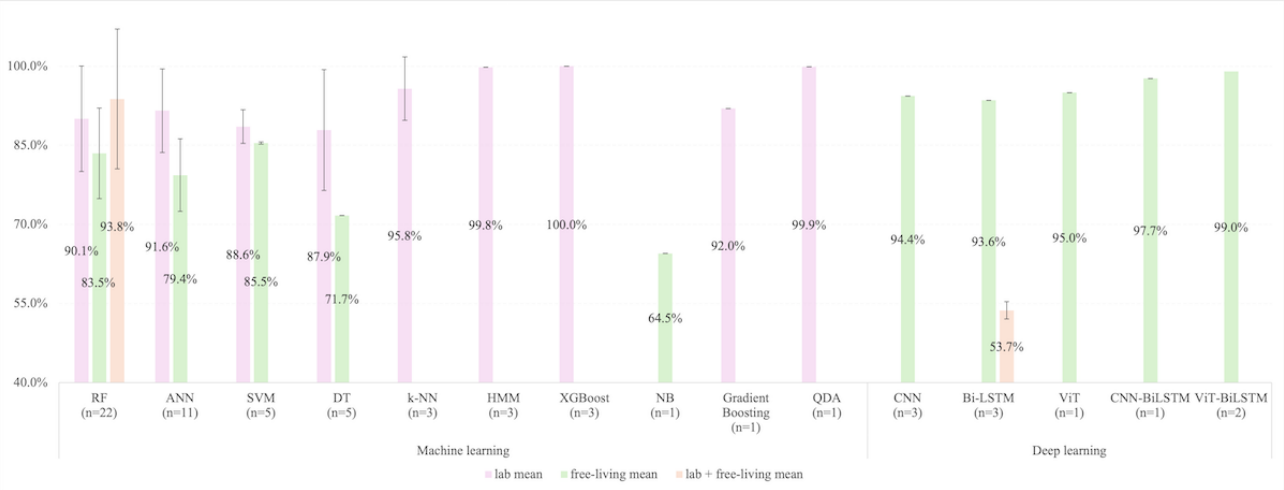
**Figure 2.** The extracted F1-score for moderate-to-vigorous physical activity from machine learning and deep learning models. Error bars, where applicable, represent SD. "n" means the number of studies using the model. ANN: artificial neural network; BiLSTM: bidirectional long short-term memory; CNN: convolutional neural network; CNN-BiLSTM: convolutional neural network and bidirectional long short-term memory; DT: decision tree; HMM: hidden Markov model; k-NN: k-nearest neighbor; LASSO: least absolute shrinkage and selection operator; QDA: quadratic discriminant analysis; RF: random forest; SVM: support vector machine; ViT: vision transformer; ViT-BiLSTM: vision transformer bidirectional long short-term memory; XGBoost: extreme gradient boosting.




However, performance disparities between lab-controlled and free-living environments underscored inherent limitations. RF models, for instance, exhibited a decline of 8.0% in $F_1$-score (88.8% lab vs 80.8 free-living) and 6.6% in accuracy (90.1% lab vs 83.5% free-living), attributed to over-reliance on static features (eg, variance, spectral entropy) that fail to generalize to unstructured movement patterns [27]. Similarly, ANN experienced reduced accuracy in free-living contexts ($F_1$-score 88.7% lab vs 75.4% free-living; accuracy 91.6% lab vs 79.4% free-living), highlighting sensitivity to signal variability introduced by nonexercise movements (eg, gesturing, device placement) [27]. The rest of the algorithms had no free-living validation.

Figures 2 and 3 stratify $F_1$-score and accuracy by model type, revealing that simpler algorithms like k-nearest neighbor and quadratic discriminant analysis achieved near-perfect lab performance (96.4%-100%) but performed less well in free-living validations.

**Figure 3.** The extracted accuracy for moderate-to-vigorous physical activity from machine learning and deep learning models. Error bars, where applicable, represent SDs. "n" means the number of studies using the model. ANN: artificial neural network; BiLSTM: bidirectional long short-term memory; CNN: convolutional neural network; CNN-BiLSTM: convolutional neural network and bidirectional long short-term memory; DT: decision tree; HMM: hidden Markov model; k-NN: k-nearest neighbor; NB: naive Bayes; QDA: quadratic discriminant analysis; RF: random forest; SVM: support vector machine; ViT: vision transformer; ViT-BiLSTM: vision transformer bidirectional long short-term memory; XGBoost: extreme gradient boosting.




### *Deep Learning*

DL architectures revolutionized MVPA detection by automating hierarchical feature extraction from raw accelerometer signals, circumventing the manual feature selection bottleneck. CNN was used in 3 studies (3/40 studies, 7.5%), achieving a mean $F_1$-score of 71.9% (shown in Figure 2) and a mean accuracy of 94.4% (shown in Figure 3) in free-living conditions [33,34,68]. Their layered structure, comprising convolutional filters (64, 128, 256, and 512 filters), pooling layers, and activation functions (such as rectified linear unit), enables granular analysis of signal dynamics.



XSL•FO
RenderX

Recurrent architectures, notably LSTM and BiLSTM, addressed temporal complexity in sustained MVPA bouts (eg, 10 min). BiLSTM, which processes sequences bidirectionally, achieves an average of $F_1$-score 71.9% and accuracy 80.3% (shown in Figures 2 and 3) in 3 studies (of 40 studies, 7.5%) (2 in free-living settings, and 1 combining both lab and free-living settings) by modeling contextual transitions (eg, walking-to-jogging) [31,33,79]. Transformers, though less prevalent, demonstrated promise in capturing long-range dependencies through self-attention mechanisms. When hybridized with vision, vision transformer (ViT), the accuracy of detecting MVPA is 95.0% ($F_1$-score 79.8%) in a free-living validating setting in 1 out of 40 studies (2.5%) [33].

Hybrid models (eg, convolutional neural network and bidirectional long short-term memory [CNN-BiLSTM], vision transformer bidirectional long short-term memory), in 2 out of 40 studies (5%), synergized spatial and temporal learning, achieving peak $F_1$-score (95.5%) [34] and peak accuracy (98.4%) in free-living settings [33].

## Task-Specific Insights

### *Classification*

Classification tasks in accelerometer-based MVPA research involve assigning discrete intensity classes, such as sedentary, light, moderate, or vigorous activity. Early methodologies predominantly used traditional ML algorithms, such as RF and SVM, which relied on handcrafted features like signal variance, spectral entropy, and movement counts. For instance, RF achieved robust performance in laboratory settings ($F_1$-score mean 91.9%, Table 3) by aggregating predictions from DTs trained on bootstrapped subsets of ActiGraph data. However, these models struggled with ambiguity in free-living environments, particularly in distinguishing light-intensity activities (eg, slow walking at 2.5 METs) from MVPA (≥3 METs). SVMs with radial basis function kernels, while effective in lab-annotated running protocols ($F_1$-score 75.4%) [28], misclassified 9.0% of slow walking (or stroll) bouts as MVPA in unstructured settings due to overlapping signal patterns [50].

**Table 3.** Taxonomy of machine learning and deep learning technologies for moderate-to-vigorous physical activity (MVPA) detection, categorized by learning paradigm.

| Learning paradigm | Key features | Algorithms | Strengths | Limitations | References |
|---|---|---|---|---|---|
| Supervised | Require labeled data (activity intensity labels) | RF[a], ANN[b], SVM[c], DT[d], XGBoost[e], HMM[f], QDA[g], LASSO[h], k-NN[i], and gradient boosting | • High accuracy with sufficient labeled data<br>• Interpretable feature importance, Robust to noise and nonlinear patterns | • Dependency on large, labeled datasets<br>• Overfitting risk<br>• Poor generalization to free-living environments | [28,29,33,36-38,50-60, 62,64-67,69-72,74-78,80] |
| Unsupervised | Work with unlabeled data, focus on clustering or feature learning | k-means, SOM[j], and autoencoders | • No need for labeled data<br>• Identifies hidden patterns in raw signals<br>• Reduces dimensionality | • Limited direct applicability to MVPA classification<br>• Lower accuracy for intensity-specific tasks<br>• Interpretability challenges | [19,29,63] |
| Hybrid | Combine supervised and unsupervised components, integrates multiple architectures | CNN-BiLSTM[k], ViT-BiLSTM[l], DLEN[m], and multitask learning frameworks | • Capture spatial and temporal dependencies<br>• Improve generalizability<br>• State-of-the-art performance in free-living settings | • High computational complexity<br>• Require large datasets<br>• Synchronization challenges for multisensor data | [29,31,33,34,68,79] |

[a]RF: random forest.

[b]ANN: artificial neural network.

[c]SVM: support vector machine.

[d]DT: decision tree.

[e]XGBoost: extreme gradient boosting.

[f]HMM: hidden Markov model.

[g]QDA: quadratic discriminant analysis.

[h]LASSO: least absolute shrinkage and selection operator.

[i]k-NN: k-nearest neighbor.

[j]SOM: self-organizing maps.

[k]CNN-BiLSTM: convolutional neural network and bidirectional long short-term memory.

[l]ViT-BiLSTM: vision transformer bidirectional long short-term memory.

[m]DLEN: deep learning ensemble network.



XSL•FO
RenderX

DL architectures, particularly CNN, addressed these limitations by automating hierarchical feature extraction directly from raw accelerometer signals. By applying convolutional filters to sliding windows of raw data, CNN detected local biomechanical patterns (eg, stride frequency during running), achieving parity with traditional hip-based cut point methods in MVPA detection by the wrist-model [68]. Subsequent advancements, such as Transformer architectures, further improved classification accuracy by using self-attention mechanisms to model long-range dependencies, outperforming CNN by 9.5% in free-living scenarios [33].

### *Estimation*

Estimation tasks focus on predicting energy expenditure metrics (eg, METs) through regression-based models to map accelerometer signals to continuous outcomes. Conventional approaches, such as linear regression-derived cut points (eg, Freedson equations), exhibited significant limitations due to oversimplified assumptions about the relationship between acceleration signals and MET values, especially for free-living activity [65,81]. ML models, such as an additive regression tree, lowered the standard error of estimation by 0.33-22.11 in lab settings using ActiGraph data [69].

DL architectures, like BiLSTM, elevated estimation accuracy by capturing temporal dependencies in accelerometer signals (eg, MET fluctuations during exercise recovery). BiLSTM achieved a mean absolute error of 0.757, with LSTM as the baseline method [79].

### *Deep Learning*

DL frameworks bridge the gap between classification and estimation by unifying feature extraction and task-specific learning with end-to-end frameworks. Multitask architectures, such as AccNet24, integrate BiLSTM layers for activity intensity with fully connected layers for MET prediction, achieving 97.7% accuracy in MVPA detection in free-living settings [31].

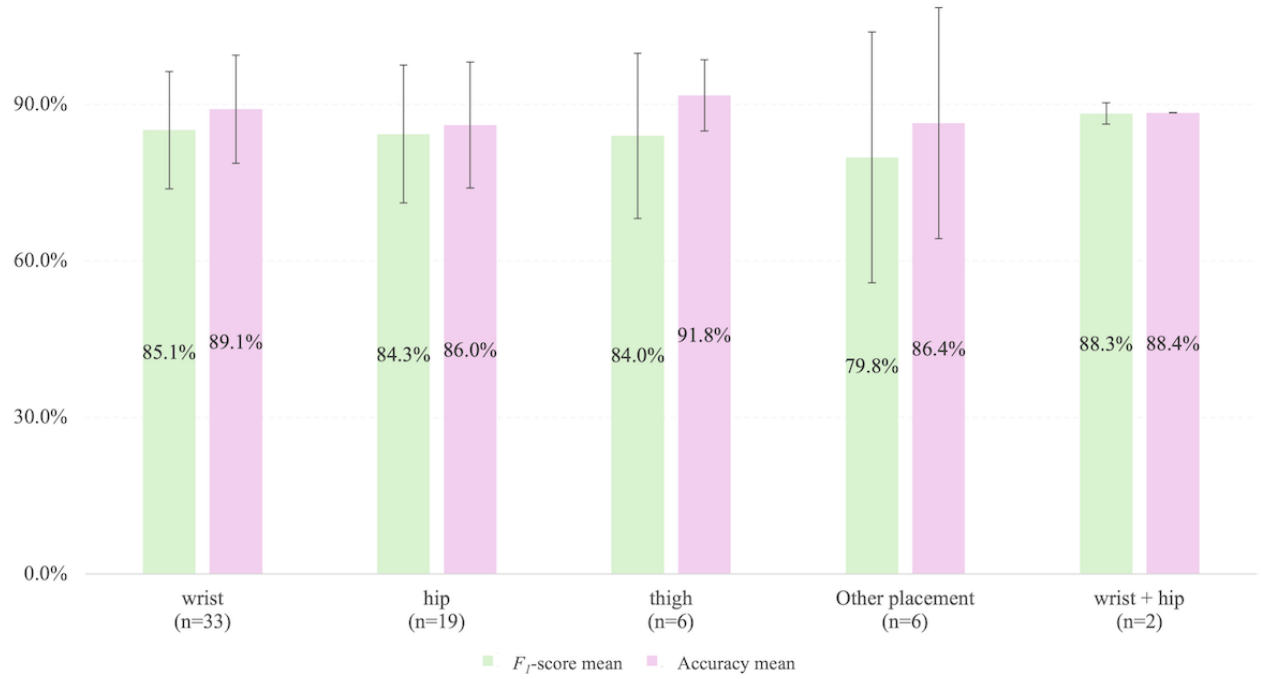
ViTs further optimized task performance via attention mechanisms that dynamically prioritized critical signal regions. For example, ViT allocated closer attention to peak acceleration intervals during jumping, outperforming Bi-LSTM by 6.2% ($F_1$-score) in free-living MVPA detection [33]. However, these advancements come with trade-offs; hybrid CNN-BiLSTM models require much more training time than traditional RF, limiting real-time deployment on wearables [31].

## Sensor Performance

### *Sensor Placement*

The efficacy of accelerometer-based MVPA estimation is significantly influenced by sensor placement. Figure 4 illustrates the averaged performance metrics ($F_1$-scores and accuracy) across sensor placements.

**Figure 4.** Mean value of F1-score and accuracy in relation to sensor placement across all validation settings. Error bars, where applicable, represent SDs. Other placements included chest, lower back, backpack, pocket, hand, and upper arm. “wrist + hip” means the multi-sensor configuration using both wrist and hip placements.




Regarding sensors placement, wrist-worn devices dominated in 75% of studies (30/40 studies), followed by hip (19/40, 47.5%), thigh (6/40, 15%), chest (1/40, 2.5%), lower back (1/40, 2.5%), backpack (1/40, 2.5%), pocket (1/40, 2.5%), hand (1/40, 2.5%), upper arm (1/40, 2.5%). Multisensor configurations (eg, wrist + hip) were applied in only 5% (2/40) of studies [28,51]; only the combination of wrist and hip was applied.

Comparative analyses on MVPA revealed that wrist-, hip-, and thigh-worn sensors exhibited comparable mean performance metrics across all validation settings (mean $F_1$-scores 85.1% vs 84.3% vs 84%, accuracy mean 89.1% vs 86% vs 91.8%) as well



XSL•FO
RenderX

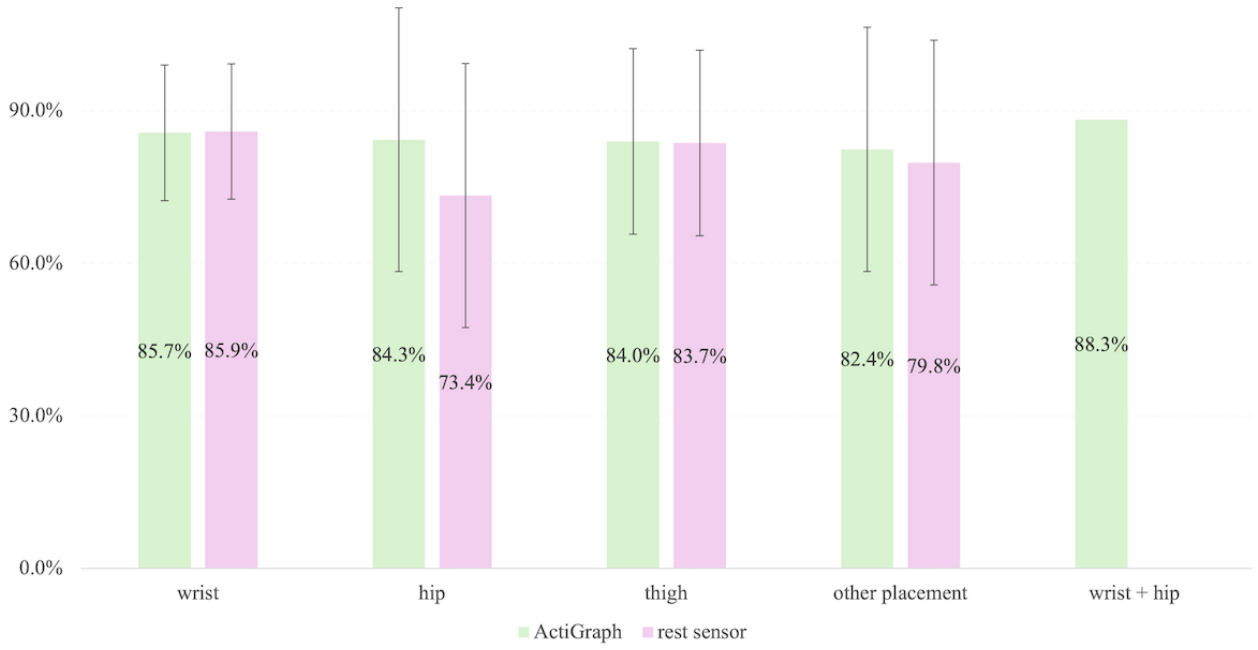
as in laboratory-controlled settings ($F_1$-score mean 88.6% vs 86.6% vs 86.5%, accuracy mean 91.7% vs 89.6% vs 97.3%). However, disparities emerged in free-living environments. While wrist- and hip-worn sensors demonstrated similar $F_1$-scores (80.3% vs 79.0%), wrist-worn devices achieved superior accuracy (86.3% vs 70.8%). This discrepancy may stem from the wrist's ability to capture a broader range of upper-body movements associated with MVPA in unstructured environments, such as arm swings during brisk walking or lifting activities, which are less pronounced in hip-worn sensors.

Notably, multisensor configurations (eg, wrist + hip) achieved the highest performance ($F_1$-score 88.3%, with 89.7% in lab and 86.8% in free-living; accuracy 88.4%, with 88.4% in lab but nonreports in free-living), bridging the gap between controlled and free-living settings. However, practical challenges, including increased participant burden due to multiple devices and synchronization complexities between heterogeneous sensors, limit their widespread adoption.

### *Sensor Type and Performance Heterogeneity*

Device specifications and sensor type further influenced model generalizability. Figure 5 shows the $F_1$-scores in relation to sensor type (ActiGraph and other types) and sensor placement. While models trained on ActiGraph data achieved a higher mean $F_1$-score (84.9% vs 83.1%), models using consumer-grade wearables surprisingly achieved a higher mean accuracy (91.8% vs 87.8%). This disparity may arise from the inherent class imbalance in free-living data, where MVPA represents a minority of activities. Accuracy can be inflated by correctly classifying the predominant sedentary and light activities, whereas the $F_1$-score provides a more balanced measure of performance specifically for the MVPA class. The higher $F_1$-score associated with ActiGraph models suggests they may be more adept at correctly identifying true MVPA bouts, which is critical for public health monitoring.

**Figure 5.** Mean value of F1-score in relation to sensor type and placement across all validation settings. Error bars, where applicable, represent SDs.




## Validation Practices

### *Ground Truth Methodologies and Their Implications*

Validation of ML and DL models for MVPA detection relies heavily on the specification of ground truth, with IC and DO predominating. IC (21/40 studies, 52.5%), considered the gold standard for energy expenditure measurement, provides MET values through oxygen consumption analysis, enabling precise alignment of accelerometer signals with intensity thresholds (eg, ≥3 METs) [19,28,37,53-56,58,59,61,62,64,66,69,70,72-74,77-79]. However, its laboratory-bound nature limits ecological validity, as structured protocols often fail to replicate free-living movement variability.

In contrast, DO (16/45 analyses, 35.6%) offers real-world applicability by annotating activities in a naturalistic setting but introduces subjectivity, particularly in distinguishing borderline intensities with 2 stages. At the first stage, participant movements were categorized into activity type (eg, sedentary, standing utilitarian tasks, walking, and running) using recordings [28,38,50,51,60,62,65,67,78] and time-stamped images from wearable cameras (eg, combining the usage of a diary in Capture-24) [31,33,53,57,75,76]. At the second stage, physical intensity was coded using references, mainly the Compendium of Physical Activities (sedentary, light, moderate, and vigorous) and Children's Activity Rating Scale (5 categories from stationary/motionless to fast translocation).

Additionally, reliance on hip-reference cut points as proxies for ground truth (2/45 analyses, 4.4%) perpetuates circular validation, wherein models trained on threshold-based labels inherit the biases of traditional regression methods [63,68]. A total of 15.6% (7/45) of analyses used a predefined activity

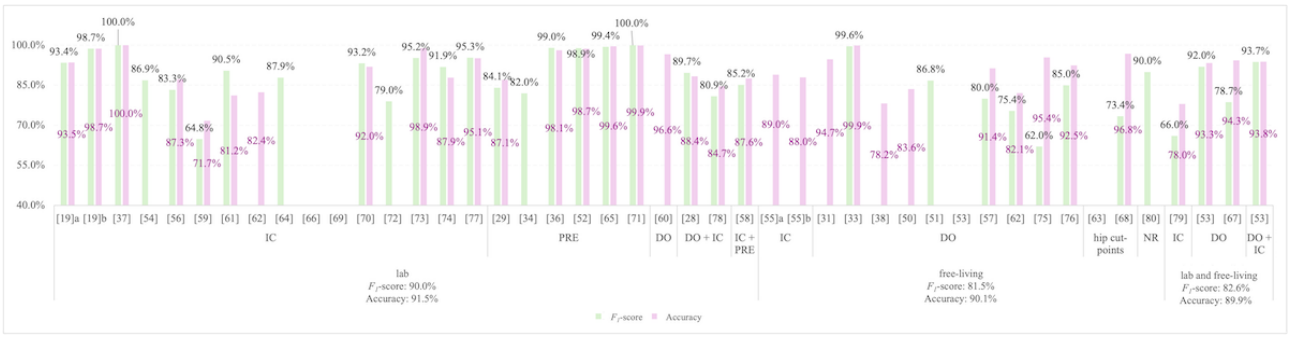


XSL•FO
RenderX

schedule in the validation process to define the ground truth [29,36,52,58,71]. Only 11.1% (5/45) of analyses used combined ground truth approaches (eg, IC + DO) [28,53,78], despite evidence that combined methods improve $F_1$-score by 8.3%-27.7% in free-living compared with IC or DO [53].

Figure 6 illustrates the distribution of ground truth methods across studies, stratified by validation setting. Lab-based studies disproportionately favored IC (19/26 analyses, 73%), while free-living validations leaned on DO (10/15 analyses, 66.7%), with lab validations outperforming free-living ($F_1$-score mean difference: 8.5%, accuracy mean difference: 1.4%).

**Figure 6.** F1-score and accuracy metrics for moderate-to-vigorous physical activity (MVPA) classification across studies, stratified by ground truth methods and validation settings. [19]a represents the MVPA classification conducted based on the 10s and 60s window lengths, while [19]b only on the 60s window length. [55]a represents the MVPA classification conducted among the children and adolescents with a mean age of 12.3 (SD 1.0) years, while [55]b represents the MVPA classification conducted among adults aged 24.9 (SD 2.6) years. The mean value of F1-score and accuracy in each group (lab, free-living, and lab and free-living) was shown underneath their names [19,28,29,31,33,34,36-38,50-80]. DO: direct observation; hip: hip reference cut-points; IC: indirect calorimetry; NR: not reported; PRE: predefine activity schedule.




### *Cross-Validation Protocols and Performance Metrics*

The trained models typically undergo rigorous evaluation during the model evaluation and validation phase to verify their generalizability and practical applicability. This process is essential to systematically assess classification accuracy across activity intensities and validate reliability under diverse user scenarios. To mitigate overfitting and ensure model robustness, *k*-fold cross-validation (11/45 analyses, 24.4%) was most commonly implemented. It partitions the dataset into *k* equally sized subsets, iteratively designating one subset as the validation set and the remaining *k*-1 subsets for training. The process is repeated *k* times to ensure all data points contribute to both training and validation. Common configurations include 10-fold and *5*-fold cross-validation, which enhance model generalizability by reducing sensitivity to specific training instances. By systematically evaluating performance across varied data partitions, this method strengthens the activity intensity classification system's reliability and mitigates overfitting, a phenomenon where models memorize training data artifacts rather than learning generalizable patterns.

Leave-one-out cross-validation, which iteratively holds out each individual data point as a test set to evaluate model performance, was also used in 13.3% of analyses (6/45 analyses). Leave-one-subject-out cross-validation, an extension of leave-one-out cross-validation designed for datasets with multiple subjects, iteratively holds out all data from one subject as the test set while training on the remaining participants. This method, used in 24.4% of analyses (11/45), was critical for assessing interindividual generalizability. Its variant, leave-10-subject-out cross-validation, appeared less frequently (1/45 analysis). In contrast, nested cross-validation, which separates hyperparameter tuning from final performance evaluation to prevent data leakage, was sparingly adopted (3/45 analyses, 6.7%) in studies.

Model efficacy was quantified using precision, recall, accuracy, and $F_1$-score, with metrics calculated iteratively to ensure objective assessment. $F_1$-score (33/45 analyses, 73.3%) and accuracy (35/45 analyses, 77.8%) were emphasized in this review, though their interpretation varied widely. Notably, studies reporting accuracy exceeding 90.0% often excluded transitional activities (eg, sit-to-stand) [19,37,65,70] or used imbalanced datasets [28,71,80], potentially inflating scores. Conversely, $F_1$-scores below 75.0% typically correlated with free-living validations, where non-MVPA movements confounded detection [62,68,75].

In the studies not reporting the metrics of MVPA detection, but including a confusion matrix, we calculated $F_1$-scores and accuracy values from the confusion matrix [28,29,33,36,52,53,57,58,61,62,67,68,70,71,73-78]. Two studies reported the $F_1$-scores of moderate physical activity (MPA) and vigorous physical activity (VPA) separately without a confusion matrix, so they averaged MPA and VPA to get MVPA [72,79].

Four studies omitted both $F_1$-score and accuracy [53,63,66,69]. For instance, Li et al [63] reported only the overall accuracy of physical activity intensity classification, while Montoye et al [66] quantified MVPA error (+1.8 min) relative to IC. Ahmadi et al [53] only provided sensitivity (MPA: 80.0%, VPA: 90.0%) and precision (MPA≈75.0%, VPA≈99.0%), and Nnamoko et al [69] reported only the standard error for estimation of personalized cut points.

## Algorithmic Bias

The performance of machine and DL models for MVPA detection is inherently tied to the physiological characteristics of the training populations. These have been mostly young, healthy adults. Persistent algorithmic bias induced by using this group can undermine the generalizability of models across the older adult and clinical cohorts.



XSL•FO
RenderX

Figures 7 and 8 stratified MVPA detection performance ($F_1$-scores and accuracy) by age group (children and adolescents, adults younger than 60 years, adults aged 60 years or older, and clinical populations) and sensor placement (wrist, hip, thigh, other). Among children and adolescents (11/40 studies, 27.5%), both $F_1$-scores (53.5%-98.9%) and accuracy (52.1%-98.7%) varied widely [19,28,29,38,50-52,55,62,63,79]. Adults under 60 years (20/40 studies, 50%) exhibited consistently high performance ($F_1$-score mean 85.8%, accuracy mean 91.7%) [31,33,34,53,55-59,61,65-68,70,71,73-76], while older adults (60 years or older) in 15% of studies (6 out of 40 studies) showed relatively reduced score ($F_1$-score mean 72.3%, accuracy mean 89.9%) [54,64,69,72,77,78]. Clinical populations (2/40 studies, 5%), on the other hand, achieved near-perfect scores ($F_1$-score 97.6%-100%, accuracy 87.9%-100%), though limited studies (n=2, one in mild cystic fibrosis, and one in type 1 diabetes) necessitate cautious interpretation [37,80].

**Figure 7.** F1-scores for moderate-to-vigorous physical activity detection across age groups and sensor placements. Brackets ("[]") represent the reference numbers; asterisks ("*") indicate lab-validated results [19,28,29,33,34,36,37,51-54,56-59,61,62,64,65,67,68,70-80]. CF: cystic fibrosis; T1D: type 1 diabetes.

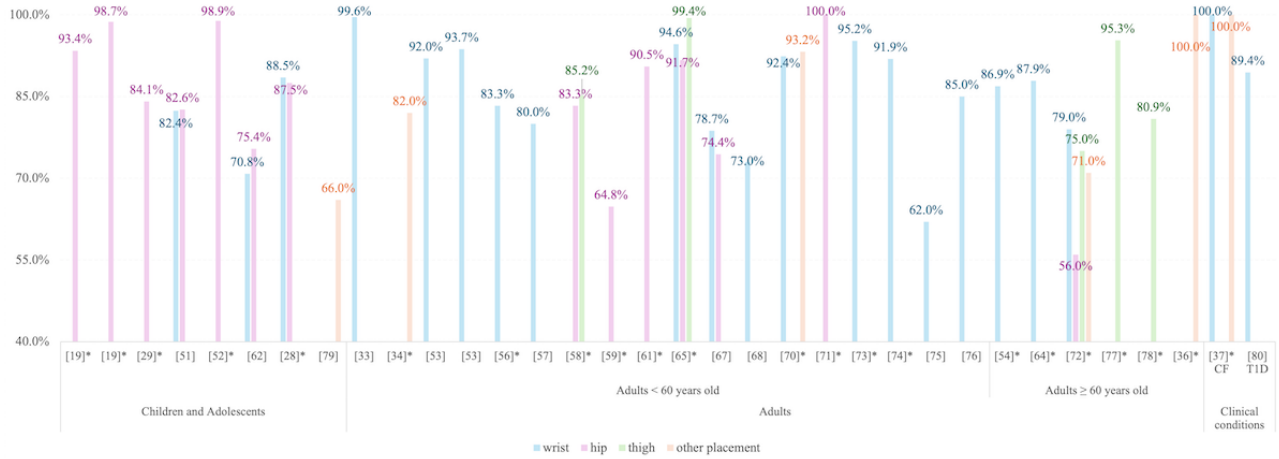
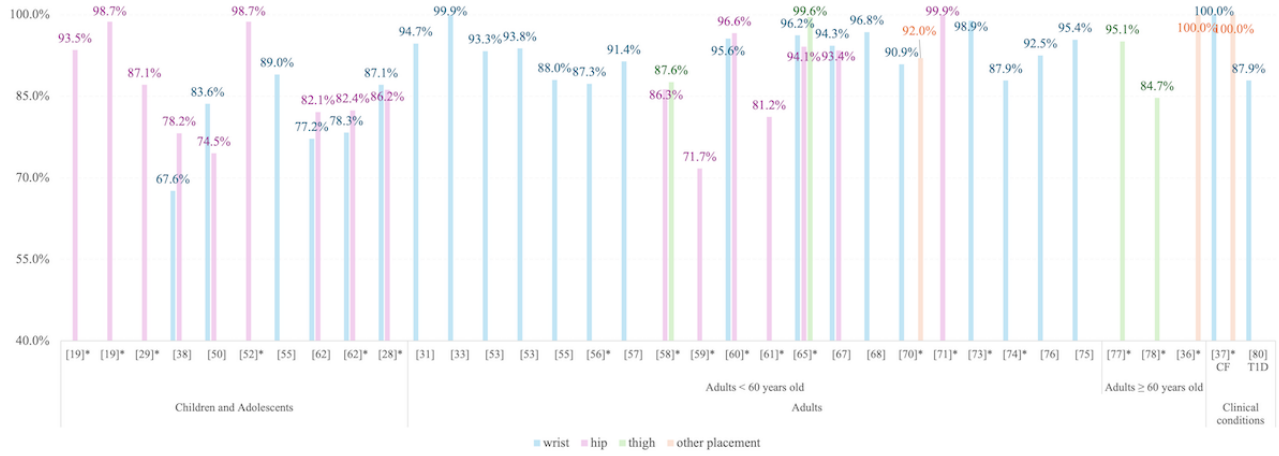



**Figure 8.** Accuracy metrics for moderate-to-vigorous physical activity detection across age groups and sensor placements. Brackets ("[]") represent the reference numbers; asterisks ("*") indicate lab-validated results [19,28,29,31,33,36-38,50,52,53,55-62,65,67,68,70,71,73-78,80]. CF: cystic fibrosis; T1D: type 1 diabetes.




## Reproducibility and Transparency Gaps

Notwithstanding the performance advancement reported, 57.5% of the studies (23/40 studies) failed to disclose code or datasets, and 60% (3/5 studies) of DL studies lacked hyperparameter specifications (eg, learning rates, batch sizes) [34,68,79]. This "black box" methodological opacity mirrors the reproducibility crisis in traditional cut point research, where proprietary algorithms replace opaque thresholds, compromising interpretability. Only 20.0% of studies (8/40) adhered to open science practices by publicly sharing the code of models [38,51,52,58,67,68,72,76], with one study providing only sample code availability [67].

The absence of standardized reporting frameworks exacerbates methodological inconsistencies. For example, window lengths for signal segmentation ranged from 1s to 60s, complicating cross-study comparisons (details shown in Table 1). While 5 studies evaluated multiple window lengths [19,29,33,51,56], optimal performance diverged across populations: 15s or 16s were preferred for adults [33,56] and preschoolers [51], whereas 60s windows were superior to 10s and 30s among preschoolers [29]. Notably, Trost et al [19] found no difference between 10s and 60s among children (mean 11, SD 2.7 years). Similarly, MET thresholds for MVPA classification varied, such as 2.8 METs [38], 3 METs [34,37,59,60,65-67,70,71,74,76,77,79], and 3.9 METs [19,29]. This variability introduces heterogeneity



XSL•FO
RenderX

in intensity categorization, undermining cross-study generalizability.

On the positive side, the proliferation of public datasets (accounting to 25%, 10/40 studies), including Capture-24 [31,33,53,57,75,76], Energy-24 [57], UOULU (University of Oulu) [60], OSU (Oregon State University) [60], the PAMAP2 Physical Activity Monitoring dataset (University of California, Irvine, UCI) [60], and the Daily and Sports Activities (the UCI Machine Learning Repository) [60] has partially mitigated by enabling benchmarking and reducing data dependency.

## Discussion

### Principal Findings

This systematic scoping review synthesizes advancements in ML and DL techniques for estimating and predicting MVPA from accelerometer data. Traditional ML models (eg, RF, ANN) demonstrated robust lab-based accuracy ($F_1$-score mean 83.6%-100%) while real-world performance declined by 8.0%-13.3% due to environmental noise and device heterogeneity. DL architectures (eg, CNN, Transformer) achieved superior performance by leveraging raw signal dynamics ($F_1$-score mean 73.6%-98.4%) in free-living settings, especially with hybrid models (CNN-BILSTM, ViT-LSTM). Wrist-worn devices were most often tested (30/40, 75.0% of studies) and performed comparably to hip/thigh placements in a lab setting ($F_1$-score mean 84.0%-84.3%, accuracy mean 86.0%-91.8%). Multiaccelerometer configurations (eg, hip + wrist) achieved the best performance (accuracy mean 88.4%) but face practical limitations. Algorithmic bias was seen to disfavor older adult participants, but not clinical populations. However, only a few studies have tested patients, limited to cystic fibrosis and type 1 diabetes.

### Methodological Advancements and Challenges

#### *From Feature Engineering to End-to-End Learning*

The evolution of MVPA detection methodologies reveals a clear paradigm shift from manual feature engineering to automated DL. Historically, ML models relied on handcrafted features (eg, spectral entropy, variance) derived from time-and frequency-domain analyses. In contrast, DL architectures, such as CNN and Transformers, automate hierarchical feature extraction from raw accelerometer signals, capturing biomechanical nuances (eg, stride variability during running) through convolutional filters and attention mechanisms [33,68]. For instance, hybrid models like CNN-BiLSTM synergized spatial and temporal learning, achieving state-of-the-art accuracy ($F_1$-score 98.4%-99.6%, accuracy 97.7%-99.0%) in free-living settings [33]. The effectiveness of this architecture is further corroborated by its successful application in related biomechanical modeling tasks, such as predicting ligament fatigue failure risk from complex signal data, highlighting its robust capability to capture critical spatiotemporal patterns [85]. Nevertheless, DL's computational intensity and reliance on high-resolution data (≥100Hz) limited deployment on resource-constrained wearables [33]. Furthermore, while DL reduced manual feature engineering burden, nearly 60.0% of models remained inaccessible due to unshared code, perpetuating reproducibility challenges (section "Reproducibility and Transparency Gaps").

Three key advancements define this evolution: (1) Static to dynamic features, unlike the fixed features of traditional ML, DL architectures dynamically extract nuanced biomechanical patterns from raw signals (section "Traditional Machine Learning" and "Deep Learning") [31,33,34,68,79]. (2) Early studies treated classification and estimation as separate tasks, but modern frameworks like AccNet24 unify these through shared neural pathways, improving efficiency (section "Task-Specific Insights") [33]. (3) Self-supervised learning, pretraining on unlabeled data, reduced annotation costs while maintaining high performance, addressing scarcity of free-living settings (section "Methodological Evolution and Comparative Insights") [33,68].

#### *Lab-to-Real-World Performance Comparison*

Although lab-validated models achieved high performance (eg, 87.9%-100% accuracy across ML techniques, section "Evolution of Feature Engineering and Model Architectures"), free-living performance experienced unstructured movement patterns and environment noise. For example, RF accuracy dropped from 90.1% (lab) to 83.5% (free-living), while wrist-based models exhibited superior adaptability to upper-body movements (eg, arm swings) in unstructured settings (section "Sensor Placement"). Notably, only 42.2% of studies validated models in real-world environments (most after 2020), highlighting a critical translational gap.

Two key insights emerge from a lab-to-real-world comparison. (1) There is a 3.1%-16.2% accuracy decline when using ML techniques (section "Evolution of Feature Engineering and Model Architectures"). Context-aware architecture, DL architectures, such as transformers, partially mitigated performance declines by leveraging context-aware attention to movement sequences (eg, detecting walking interruptions), achieving accuracy of 95.0% in free-living scenarios [33]. (2) There is an algorithmic bias across age groups that hinders real-world deployment (section "Algorithmic Bias").

#### *Validation and Reproducibility*

A key challenge lies in inconsistent validation protocols. While IC provided precise MET-based thresholds, its lab-bound nature limited ecological validity [58]. Conversely, DO offered real-world applicability but introduced subjectivity in intensity classification [86,87]. Moreover, disparities in metrics reporting (eg, exclusion of transitional activities) [19,37,65,70] and variable parameters (eg, MET threshold: 2.8-3.9, window lengths 1-60s) hindered cross-study comparability (section "Reproducibility and Transparency Gaps"). Compounding these issues, 42.5% of studies adhered to open science practices, perpetuating a "new cut-point conundrum" akin to proprietary regression thresholds.

Our synthesis reveals a vicious cycle underpinning the translational challenges in AI-driven MVPA monitoring. The foundational issue is the lack of standardized validation protocols. Inconsistent MET thresholds and variable data window lengths mean that models are trained and evaluated on



XSL•FO
RenderX

fundamentally different definitions of MVPA. This directly contributes to the lab-to-real-world performance gap, as a model calibrated with one protocol fails to generalize to data collected under another. Furthermore, this inconsistency, when combined with the prevalent lack of code sharing, makes it impossible to audit, replicate, or fairly compare models. Consequently, this opacity hinders the identification and correction of algorithmic bias against underrepresented populations, as the root cause of poor performance, a flawed model versus an incompatible validation method, cannot be discerned. Thus, these challenges are not isolated but are synergistic barriers that collectively impede the development of truly generalizable and equitable models.

### *Sensor Performance and Device Bias*

Device placement and type emerged as critical determinants of model performance, as evidenced in the section "Sensor Performance." For instance, while ActiGraph-trained models achieved high lab accuracy ($F_1$-score 79.9%, accuracy 90.5%), they underperformed on consumer wearables (eg, Samsung smartwatch, $F_1$-score mean difference 3.2%) due to differences in sensor calibration and sampling rates (Table 1 and Figure 5) [56]. Additionally, interdevice variability across brands (eg, Axivity vs GENEActiv) exacerbated performance inconsistencies, particularly in free-living settings. Notably, optimal sensor placement (eg, wrist vs hip) influenced adaptability to movement patterns, with wrist-worn devices showing superior capture of upper-body dynamics (eg, arm swings) but struggling with lower-body activities [40,88]. These findings highlight the need for device-agnostic training pipelines to mitigate performance variability across brands and placements.

## Translational Opportunities and Challenges

### *Public Health and Clinical Integration*

Wrist-worn devices demonstrated comparable accuracy to hip/thigh placements in lab settings ($F_1$-score 84.0%-84.3%, accuracy 86.0%-91.8%) and superior adaptability to free-living upper-body movements (Figure 4), supporting their feasibility for scalable monitoring. However, ActiGraph's dominance (n=30, 75.0% of studies) and limited validation on consumer wearables (eg, smartwatches) hinder real-world applicability. Clinically, models achieved high accuracy in controlled settings for cystic fibrosis and type 1 diabetes, but small sample sizes and structured protocols limit ecological validity [37,80]. Expanding validation studies to more diverse clinical populations (eg, mobility impairments) is critical.

### *Age and Population Disparities*

Results revealed systemic biases across the age range (section "Algorithmic Bias"). For instance, models trained on adults misclassified MVPA in children ($F_1$-score mean difference: –5.7%) due to developmental differences in stride length and metabolic variability [19]. Studies involving preschoolers reported accuracy fluctuations between 53.7% and 88.4%, reflecting challenges in modeling erratic movement patterns typical of young children [28,29,38,50,52,63,79]. Conversely, older adults (60 years or older) exhibited reduced accuracy ($F_1$-score mean 77.9%) due to slower gait speeds, postural instability, and comorbidities that alter movement signatures [54,64,69,72,77,78]. Wrist-based model, for example, underestimated MVPA in this cohort by 6.0%-16.6% compared with thigh-worn sensors, highlighting the need for age-specific calibration [78].

## Emerging Innovations

Hybrid DL models have emerged as a powerful approach. For instance, integrating LSTM with CNN (CNN-LSTM) or ViTs (vision transformer bidirectional long short-term memory) enables the capture of spatial-temporal patterns in accelerometry data [33,34]. Building on this, BiLSTM layers further enhance temporal dependency modeling by analyzing sequences in both forward and backward directions [31].

In parallel, image-based feature extraction methods, such as converting raw accelerometer signals into Gramian angular field images, have improved feature learning by transforming time-series data into visual representations [33]. Additionally, multisensor fusion strategies—combining data from hip, wrist, and thigh placements—address variability in sensor positioning, boosting model robustness [28,34]. Furthermore, transfer learning leverages pretrained architectures like ResNet101, adapting them for accelerometer classification tasks [31].

Another key innovation lies in advanced feature engineering. Autonomous feature extraction via CNN reduces reliance on handcrafted features [68], while time-frequency domain fusion (eg, spectral power) enhances activity discrimination [52]. Notably, real-time and edge computing advancements explore lightweight models through pruning and quantization, enabling deployment on wearable devices [33].

However, significant challenges remain. First, models trained in controlled lab settings often generalize poorly to free-living environments due to uncontrolled variability [52]. Moreover, short and heterogeneous activity bouts, common in populations like preschoolers, result in mixed-activity windows that complicate classification [29]. Another critical challenge is sensor placement variability, as signal patterns differ across body positions [37]. Compounding this, class imbalance from overrepresented sedentary/light activities skews model performance [75]. Additionally, computational complexity limits real-time use, as seen in resource-heavy models like AccNet24 [31]. Finally, distinguishing biomechanically similar activities (eg, climbing vs walking) remains problematic [29].

## Future Directions

To address existing gaps, 4 interconnected priorities emerge. First, resolving inconsistencies in ground truth methods, such as variable MET thresholds (2.8-3.9 METs) and window lengths (1-60s), is critical. This requires standardized validation frameworks, including consensus guidelines and open datasets (eg, Capture-24), to harmonize protocols and reduce discrepancies in intensity classification [64].

Second, prioritizing free-living validation is essential to bridge the lab-to-real-world performance gap. For instance, RF models exhibit accuracy declines from 90.1% in lab settings to 83.5% in free-living environments. Concurrently, diversifying training



XSL•FO
RenderX

data to include underrepresented groups, such as the older adult, pediatric, and clinical populations, will improve generalizability and mitigate age-related biases [62,72].

Third, advancing algorithmic fairness through regulatory frameworks and bias audits is imperative. This includes expanding datasets to encompass more and diverse clinical cohorts while addressing disparities in model performance across the age range. Additionally, mandating open science practices, such as code/data sharing and hyperparameter transparency, will enhance reproducibility and resolve the "new cut-point conundrum" plaguing activity intensity thresholds.

Finally, optimizing DL architectures, such as quantized models or hybrid CNN-BiLSTM frameworks, for low-power wearables will enable real-world deployment while maintaining computational efficiency [31,33].

Looking ahead, these priorities align with broader calls for standardization and interpretability. For example, improving the "black-box" nature of DL models [68] and harmonizing evaluation metrics will foster clinical trust. Moreover, lightweight, edge-compatible architectures and multimodal data integration represent promising pathways to overcome current limitations in real-world MVPA monitoring.

### Limitations and Methodological Considerations

The strengths of this review include the rigorous adherence to PRISMA-ScR guidelines, a comprehensive search strategy across 3 electronic databases (PubMed, IEEE Xplore, Web of Science, and others via manual citation tracking), and systematic screening of 1938 records. The methodology prioritized transparency through dual-reviewer full-text screening to resolve discrepancies and consultation to ensure methodological rigor. By focusing on peer-reviewed studies, we aimed to synthesize evidence grounded in empirical validation, thereby minimizing inclusion of speculative or opinion-based articles.

However, several limitations warrant consideration. First, the exclusion of gray literature (eg, unpublished trials, industry reports, or conference proceedings) may have omitted insights from ongoing or unsuccessful implementation efforts, particularly those led by technology developers or health care providers. This introduces potential publication bias, as negative results or pragmatic challenges in real-world deployment are often underrepresented in peer-reviewed journals. Second, our decision to exclude non–peer-reviewed studies and prioritize articles reporting empirical implementation in clinical or free-living settings risks overlooking formative research, such as feasibility studies or pilot trials, which could offer valuable lessons for scalable AI integration.

A further limitation arises from our emphasis on the highest reported performance metrics (eg, $F_1$-scores, accuracy) across studies. While this approach highlights peak algorithmic capabilities, it may overestimate real-world applicability, as optimal configurations (eg, 15-second windows for adults, multisensor placements) often lack generalizability to diverse populations or unstructured environments. For instance, models achieving 99.0% accuracy in lab settings may exhibit significant performance degradation in free-living contexts due to uncontrolled variables like device heterogeneity or nonexercise movements.

Methodologically, while the Arksey and O'Malley framework does not mandate quality appraisal, the inclusion of studies with heterogeneous validation protocols (eg, variable MET thresholds, ground-truth methodologies) complicates cross-study comparisons. Future reviews could strengthen synthesis by incorporating quality assessment tools to evaluate bias risk and methodological consistency. Last, the predominance of studies using young, healthy cohorts limits insights into algorithmic fairness and generalizability for older adult or clinical populations, underscoring the need for more inclusive training datasets.

These considerations do not diminish the review's contributions but highlight critical gaps, such as reproducibility challenges and translational biases, that must be addressed to advance equitable, real-world deployment of AI-driven MVPA monitoring tools.

### Conclusions

This systematic scoping review highlights that ML and DL have significantly advanced in the detection of MVPA by using accelerometer data, yet persistent gaps in generalizability and transparency hinder real-world impact. To bridge the lab-to-real-world divide, collaborative efforts across public health and computer science must prioritize reproducibility, inclusive design, and robust validation. By addressing these challenges, AI-driven tools can fulfill their potential as scalable, equitable solutions for advancing global physical activity research and intervention.

### Acknowledgments

During the preparation of this work, the authors used DeepSeek to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

### Funding

This work was supported by the Shanghai Key Lab of Human Performance (Shanghai University of Sport) (grant number 11DZ2261100). The funder provided financial support for the article processing charge of this review.



XSL•FO
RenderX

## Authors' Contributions

EJCDG and YZ contributed to the conceptualization of the study. SRBVDV and YZ were responsible for the methodology and formal analysis. PC provided resources and supervised the project, with project administration shared between PC and YZ. YZ prepared the original draft and created the visualizations, while EJCDG, PC, SRBVDV, and YZ contributed to the review and editing of the manuscript. All authors have read and approved the final version of the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 22 KB - ijmr_v15i1e76601_app1.docx ]

Multimedia Appendix 2
The full summary of included studies (N=40 studies, ranked by health condition and alphabets of author names).
[DOCX File , 83 KB - ijmr_v15i1e76601_app2.docx ]

Multimedia Appendix 3
PRISMA-ScR checklist.
[PDF File (Adobe PDF File), 177 KB - ijmr_v15i1e76601_app3.pdf ]

## References


1. Herrmann SD, Willis EA, Ainsworth BE, Barreira TV, Hastert M, Kracht CL, et al. 2024 adult compendium of physical activities: a third update of the energy costs of human activities. J Sport Health Sci 2024;13(1):6-12 [FREE Full text] [doi: 10.1016/j.jshs.2023.10.010] [Medline: 38242596]
2. Jetté M, Sidney K, Blümchen G. Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. Clin Cardiol 1990;13(8):555-565 [FREE Full text] [doi: 10.1002/clc.4960130809] [Medline: 2204507]
3. Lin H, Sardana M, Zhang Y, Liu C, Trinquart L, Benjamin EJ, et al. Association of habitual physical activity with cardiovascular disease risk. Circ Res 2020;127(10):1253-1260 [FREE Full text] [doi: 10.1161/CIRCRESAHA.120.317578] [Medline: 32842915]
4. Ried-Larsen M, Grøntved A, Kristensen PL, Froberg K, Andersen LB. Moderate-and-vigorous physical activity from adolescence to adulthood and subclinical atherosclerosis in adulthood: prospective observations from the European Youth Heart Study. Br J Sports Med 2015;49(2):107-112. [doi: 10.1136/bjsports-2013-092409] [Medline: 23584827]
5. Chen Y, Chen Z, Pan L, Ma Z, Zhang H, Li X, et al. Effect of moderate and vigorous aerobic exercise on incident diabetes in adults with obesity: a 10-year follow-up of a randomized clinical trial. JAMA Intern Med 2023;183(3):272-275 [FREE Full text] [doi: 10.1001/jamainternmed.2022.6291] [Medline: 36716009]
6. Moore SC, Patel AV, Matthews CE, Berrington de Gonzalez A, Park Y, Katki HA, et al. Leisure time physical activity of moderate to vigorous intensity and mortality: a large pooled cohort analysis. PLoS Med 2012;9(11):e1001335 [FREE Full text] [doi: 10.1371/journal.pmed.1001335] [Medline: 23139642]
7. World Health Organization. WHO Guidelines on Physical Activity and Sedentary Behaviour. Geneva: World Health Organization; 2020.
8. World Health Organization. Guidelines on Physical Activity, Sedentary Behaviour and Sleep for Children under 5 Years of Age. Geneva: World Health Organization; 2019.
9. Cliff DP, Reilly JJ, Okely AD. Methodological considerations in using accelerometers to assess habitual physical activity in children aged 0-5 years. J Sci Med Sport 2009;12(5):557-567. [doi: 10.1016/j.jsams.2008.10.008] [Medline: 19147404]
10. Cain KL, Sallis JF, Conway TL, Van Dyck D, Calhoon L. Using accelerometers in youth physical activity studies: a review of methods. J Phys Act Health 2013;10(3):437-450. [doi: 10.1123/jpah.10.3.437] [Medline: 23620392]
11. Copeland JL, Esliger DW. Accelerometer assessment of physical activity in active, healthy older adults. J Aging Phys Act 2009;17(1):17-30 [FREE Full text] [doi: 10.1123/japa.17.1.17] [Medline: 19299836]
12. Strath SJ, Pfeiffer KA, Whitt-Glover MC. Accelerometer use with children, older adults, and adults with functional limitations. Med Sci Sports Exerc 2012;44(1 Suppl 1):S77-S85 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399eb1] [Medline: 22157778]
13. Jago R, Zakeri I, Baranowski T, Watson K. Decision boundaries and receiver operating characteristic curves: new methods for determining accelerometer cutpoints. J Sports Sci 2007;25(8):937-944. [doi: 10.1080/02640410600908027] [Medline: 17474047]
14. Welk GJ. Principles of design and analyses for the calibration of accelerometry-based activity monitors. Med Sci Sports Exerc 2005;37(11 Suppl):S501-S511. [doi: 10.1249/01.mss.0000185660.38335.de] [Medline: 16294113]




XSL•FO
RenderX

15. Freedson PS, Melanson E, Sirard J. Calibration of the computer science and applications, inc. accelerometer. Med Sci Sports Exerc 1998;30(5):777-781. [doi: 10.1097/00005768-199805000-00021] [Medline: 9588623]
16. Welch WA, Bassett DR, Thompson DL, Freedson PS, Staudenmayer JW, John D, et al. Classification accuracy of the wrist-worn gravity estimator of normal everyday activity accelerometer. Med Sci Sports Exerc 2013;45(10):2012-2019 [FREE Full text] [doi: 10.1249/MSS.0b013e3182965249] [Medline: 23584403]
17. Janssen X, Cliff DP, Reilly JJ, Hinkley T, Jones RA, Batterham M, et al. Predictive validity and classification accuracy of ActiGraph energy expenditure equations and cut-points in young children. PLoS One 2013;8(11):e79124 [FREE Full text] [doi: 10.1371/journal.pone.0079124] [Medline: 24244433]
18. Trost SG, Way R, Okely AD. Predictive validity of three ActiGraph energy expenditure equations for children. Med Sci Sports Exerc 2006;38(2):380-387. [doi: 10.1249/01.mss.0000183848.25845.e0] [Medline: 16531910]
19. Trost SG, Wong W, Pfeiffer KA, Zheng Y. Artificial neural networks to predict activity type and energy expenditure in youth. Med Sci Sports Exerc 2012;44(9):1801-1809 [FREE Full text] [doi: 10.1249/MSS.0b013e318258ac11] [Medline: 22525766]
20. Trost SG, Loprinzi PD, Moore R, Pfeiffer KA. Comparison of accelerometer cut points for predicting activity intensity in youth. Med Sci Sports Exerc 2011;43(7):1360-1368. [doi: 10.1249/MSS.0b013e318206476e] [Medline: 21131873]
21. Ellis K, Kerr J, Godbole S, Staudenmayer J, Lanckriet G. Hip and wrist accelerometer algorithms for free-living behavior classification. Med Sci Sports Exerc 2016;48(5):933-940 [FREE Full text] [doi: 10.1249/MSS.0000000000000840] [Medline: 26673126]
22. Trost SG, Brookes DSK, Ahmadi MN. Evaluation of wrist accelerometer cut-points for classifying physical activity intensity in youth. Front Digit Health 2022;4:884307 [FREE Full text] [doi: 10.3389/fdgth.2022.884307] [Medline: 35585912]
23. Bassett DR, Rowlands A, Trost SG. Calibration and validation of wearable monitors. Med Sci Sports Exerc 2012;44(1 Suppl 1):S32-S38 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399cf7] [Medline: 22157772]
24. Wang H, Ma C, Zhou L. A brief review of machine learning and its application. 2009 Presented at: International Conference on Information Engineering and Computer Science; 2009 December 19-20; Wuhan, China p. 19-20. [doi: 10.1109/iciecs.2009.5362936]
25. Mjolsness E, DeCoste D. Machine learning for science: state of the art and future prospects. Science 2001;293(5537):2051-2055. [doi: 10.1126/science.293.5537.2051] [Medline: 11557883]
26. Ruch N, Joss F, Jimmy G, Melzer K, Hänggi J, Mäder U. Neural network versus activity-specific prediction equations for energy expenditure estimation in children. J Appl Physiol (1985) 2013;115(9):1229-1236 [FREE Full text] [doi: 10.1152/japplphysiol.01443.2012] [Medline: 23990244]
27. Farrahi V, Niemelä M, Kangas M, Korpelainen R, Jämsä T. Calibration and validation of accelerometer-based activity monitors: a systematic review of machine-learning approaches. Gait Posture 2019;68:285-299 [FREE Full text] [doi: 10.1016/j.gaitpost.2018.12.003] [Medline: 30579037]
28. Trost SG, Cliff DP, Ahmadi MN, Tuc NV, Hagenbuchner M. Sensor-enabled activity class recognition in preschoolers: hip versus wrist data. Med Sci Sports Exerc 2018;50(3):634-641. [doi: 10.1249/MSS.0000000000001460] [Medline: 29059107]
29. Hagenbuchner M, Cliff DP, Trost SG, Van Tuc N, Peoples GE. Prediction of activity type in preschool children using machine learning techniques. J Sci Med Sport 2015;18(4):426-431. [doi: 10.1016/j.jsams.2014.06.003] [Medline: 25088983]
30. Alinia P, Saeedi R, Fallahzadeh R, Rokni A, Ghasemzadeh H. A reliable and reconfigurable signal processing framework for estimation of metabolic equivalent of task in wearable sensors. IEEE J Sel Top Signal Process 2016;10(5):842-853. [doi: 10.1109/jstsp.2016.2569472]
31. Farrahi V, Muhammad U, Rostami M, Oussalah M. AccNet24: a deep learning framework for classifying 24-hour activity behaviours from wrist-worn accelerometer data under free-living environments. Int J Med Inform 2023;172:105004 [FREE Full text] [doi: 10.1016/j.ijmedinf.2023.105004] [Medline: 36724729]
32. Ng JYY, Zhang JH, Hui SS, Jiang G, Yau F, Cheng J, et al. Development of a multi-wear-site, deep learning-based physical activity intensity classification algorithm using raw acceleration data. PLoS One 2024;19(3):e0299295 [FREE Full text] [doi: 10.1371/journal.pone.0299295] [Medline: 38452147]
33. Wang L, Luo Z, Zhang T. A novel ViT-BILSTM model for physical activity intensity classification in adults using gravity-based acceleration. BMC Biomed Eng 2025;7(1):2. [doi: 10.1186/s42490-025-00088-2] [Medline: 39891283]
34. Barua A, Jiang X, Fuller D. The effectiveness of simple heuristic features in sensor orientation and placement problems in human activity recognition using a single smartphone accelerometer. Biomed Eng Online 2024;23(1):21 [FREE Full text] [doi: 10.1186/s12938-024-01213-3] [Medline: 38368358]
35. Pfeiffer KA, Clevenger KA, Kaplan A, Van Camp CA, Strath SJ, Montoye AHK. Accessibility and use of novel methods for predicting physical activity and energy expenditure using accelerometry: a scoping review. Physiol Meas 2022;43(9). [doi: 10.1088/1361-6579/ac89ca] [Medline: 35970175]
36. Andò B, Manenti M, Greco D, Pistorio A. An embedded sensing methodology for the classification of activity rate. 2024 Presented at: IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE); 2024 October 21-23; St Albans, United Kingdom p. 21-23. [doi: 10.1109/metroxraine62247.2024.10796795]




XSL•FO
RenderX

37. Bianchim MS, McNarry MA, Barker AR, Williams CA, Denford S, Thia L, et al. A machine learning approach for physical activity recognition in cystic fibrosis. Meas Phys Educ Exerc Sci 2023;28(2):172-181. [doi: 10.1080/1091367x.2023.2271444]
38. Ahmadi MN, Trost SG. Device-based measurement of physical activity in pre-schoolers: comparison of machine learning and cut point methods. PLoS One 2022;17(4):e0266970 [FREE Full text] [doi: 10.1371/journal.pone.0266970] [Medline: 35417492]
39. Clevenger KA, Montoye AHK, Van Camp CA, Strath SJ, Pfeiffer KA. Methods for estimating physical activity and energy expenditure using raw accelerometry data or novel analytical approaches: a repository, framework, and reporting guidelines. Physiol Meas 2022;43(9). [doi: 10.1088/1361-6579/ac89c9] [Medline: 35970174]
40. Gao Z, Liu W, McDonough DJ, Zeng N, Lee JE. The dilemma of analyzing physical activity and sedentary behavior with wrist accelerometer data: challenges and opportunities. J Clin Med 2021;10(24):5951 [FREE Full text] [doi: 10.3390/jcm10245951] [Medline: 34945247]
41. Allahbakhshi H, Hinrichs T, Huang H, Weibel R. The key factors in physical activity type detection using real-life data: a systematic review. Front Physiol 2019;10:75 [FREE Full text] [doi: 10.3389/fphys.2019.00075] [Medline: 30809152]
42. Gupta N, Gupta SK, Pathak RK, Jain V, Rashidi P, Suri JS. Human activity recognition in artificial intelligence framework: a narrative review. Artif Intell Rev 2022;55(6):4755-4808 [FREE Full text] [doi: 10.1007/s10462-021-10116-x] [Medline: 35068651]
43. Attal F, Mohammed S, Dedabrishvili M, Chamroukhi F, Oukhellou L, Amirat Y. Physical human activity recognition using wearable sensors. Sensors (Basel) 2015;15(12):31314-31338 [FREE Full text] [doi: 10.3390/s151229858] [Medline: 26690450]
44. Zhang S, Li Y, Zhang S, Shahabi F, Xia S, Deng Y, et al. Deep learning in human activity recognition with wearable sensors: a review on advances. Sensors (Basel) 2022;22(4):1476 [FREE Full text] [doi: 10.3390/s22041476] [Medline: 35214377]
45. de Almeida Mendes M, da Silva ICM, Ramires VV, Reichert FF, Martins RC, Tomasi E. Calibration of raw accelerometer data to measure physical activity: a systematic review. Gait Posture 2018;61:98-110. [doi: 10.1016/j.gaitpost.2017.12.028] [Medline: 29324298]
46. Cust EE, Sweeting AJ, Ball K, Robertson S. Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. J Sports Sci 2019;37(5):568-600. [doi: 10.1080/02640414.2018.1521769] [Medline: 30307362]
47. Trujillo-Guerrero MF, Román-Niemes S, Jaén-Vargas M, Cadiz A, Fonseca R, Serrano-Olmedo JJ. Accuracy comparison of CNN, LSTM, and transformer for activity recognition using IMU and visual markers. IEEE Access 2023;11:106650-106669. [doi: 10.1109/access.2023.3318563]
48. Lavanya TV. A machine learning approach for predicting physical activity intensity from wearable sensor data. 2024 Presented at: 7th International Conference on Circuit Power and Computing Technologies (ICCPCT); 2024 August 7-8; Poruvazhy, Kerala, India p. 8-9. [doi: 10.1109/iccpct61902.2024.10673260]
49. Lavanya TV, Sivaraman K. Predictive analytics for monitoring and classifying physical activity intensity using smartphone sensors. 2024 Presented at: 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS); 2024July 10-12; Coimbatore, India. [doi: 10.1109/icscss60660.2024.10625133]
50. Ahmadi MN, Brookes D, Chowdhury A, Pavey T, Trost SG. Free-living evaluation of laboratory-based activity classifiers in preschoolers. Med Sci Sports Exerc 2020;52(5):1227-1234. [doi: 10.1249/MSS.0000000000002221] [Medline: 31764460]
51. Ahmadi MN, Pavey TG, Trost SG. Machine learning models for classifying physical activity in free-living preschool children. Sensors (Basel) 2020;20(16):4364 [FREE Full text] [doi: 10.3390/s20164364] [Medline: 32764316]
52. Ahmadi MN, Pfeiffer KA, Trost SG. Physical activity classification in youth using raw accelerometer data from the hip. Meas Phys Educ Exerc Sci 2020;24(2):129-136. [doi: 10.1080/1091367x.2020.1716768]
53. Ahmadi MN, Hamer M, Gill JMR, Murphy M, Sanders JP, Doherty A, et al. Brief bouts of device-measured intermittent lifestyle physical activity and its association with major adverse cardiovascular events and mortality in people who do not exercise: a prospective cohort study. Lancet Public Health 2023;8(10):e800-e810 [FREE Full text] [doi: 10.1016/S2468-2667(23)00183-4] [Medline: 37777289]
54. Bai C, Wanigatunga AA, Saldana S, Casanova R, Manini TM, Mardini MT. Are machine learning models on wrist accelerometry robust against differences in physical performance among older adults? Sensors (Basel) 2022;22(8):3061 [FREE Full text] [doi: 10.3390/s22083061] [Medline: 35459045]
55. Chen D, Du Y, Liu Y, Hong J, Yin X, Zhu Z, et al. Development and validation of a smartwatch algorithm for differentiating physical activity intensity in health monitoring. Sci Rep 2024;14(1):9530 [FREE Full text] [doi: 10.1038/s41598-024-59602-6] [Medline: 38664457]
56. Davoudi A, Wanigatunga AA, Kheirkhahan M, Corbett DB, Mendoza T, Battula M, et al. Accuracy of samsung gear s smartwatch for activity recognition: validation study. JMIR Mhealth Uhealth 2019;7(2):e11270 [FREE Full text] [doi: 10.2196/11270] [Medline: 30724739]
57. Doherty A, Smith-Byrne K, Ferreira T, Holmes MV, Holmes C, Pulit SL, et al. GWAS identifies 14 loci for device-measured physical activity and sleep duration. Nat Commun 2018;9(1):5257 [FREE Full text] [doi: 10.1038/s41467-018-07743-4] [Medline: 30531941]




XSL•FO
RenderX

58. Ellingson LD, Schwabacher IJ, Kim Y, Welk GJ, Cook DB. Validity of an integrative method for processing physical activity data. Med Sci Sports Exerc 2016;48(8):1629-1638. [doi: 10.1249/MSS.0000000000000915] [Medline: 27015380]
59. Ellingson LD, Hibbing PR, Kim Y, Frey-Law LA, Saint-Maurice PF, Welk GJ. Lab-based validation of different data processing methods for wrist-worn ActiGraph accelerometers in young adults. Physiol Meas 2017;38(6):1045-1060. [doi: 10.1088/1361-6579/aa6d00] [Medline: 28481750]
60. Farrahi V, Niemela M, Tjurin P, Kangas M, Korpelainen R, Jamsa T. Evaluating and enhancing the generalization performance of machine learning models for physical activity intensity prediction from raw acceleration data. IEEE J Biomed Health Inform 2020;24(1):27-38. [doi: 10.1109/JBHI.2019.2917565] [Medline: 31107668]
61. Freedson PS, Lyden K, Kozey-Keadle S, Staudenmayer J. Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: validation on an independent sample. J Appl Physiol (1985) 2011;111(6):1804-1812 [FREE Full text] [doi: 10.1152/japplphysiol.00309.2011] [Medline: 21885802]
62. Hibbing PR, Ellingson LD, Dixon PM, Welk GJ. Adapted sojourn models to estimate activity intensity in youth: a suite of tools. Med Sci Sports Exerc 2018;50(4):846-854. [doi: 10.1249/MSS.0000000000001486] [Medline: 29135657]
63. Li S, Howard JT, Sosa ET, Cordova A, Parra-Medina D, Yin Z. Calibrating wrist-worn accelerometers for physical activity assessment in preschoolers: machine learning approaches. JMIR Form Res 2020;4(8):e16727 [FREE Full text] [doi: 10.2196/16727] [Medline: 32667893]
64. Mardini MT, Bai C, Wanigatunga AA, Saldana S, Casanova R, Manini TM. Age differences in estimating physical activity by wrist accelerometry using machine learning. Sensors (Basel) 2021;21(10):3352 [FREE Full text] [doi: 10.3390/s21103352] [Medline: 34065906]
65. Montoye AHK, Pivarnik JM, Mudd LM, Biswas S, Pfeiffer KA. Validation and comparison of accelerometers worn on the hip, thigh, and wrists for measuring physical activity and sedentary behavior. AIMS Public Health 2016;3(2):298-312 [FREE Full text] [doi: 10.3934/publichealth.2016.2.298] [Medline: 29546164]
66. Montoye AHK, Pivarnik JM, Mudd LM, Biswas S, Pfeiffer KA. Evaluation of the activPAL accelerometer for physical activity and energy expenditure estimation in a semi-structured setting. J Sci Med Sport 2017;20(11):1003-1007. [doi: 10.1016/j.jsams.2017.04.011] [Medline: 28483558]
67. Montoye AHK, Westgate BS, Clevenger KA, Pfeiffer KA, Vondrasek JD, Fonley MR, et al. Individual versus group calibration of machine learning models for physical activity assessment using body-worn accelerometers. Med Sci Sports Exerc 2021;53(12):2691-2701. [doi: 10.1249/MSS.0000000000002752] [Medline: 34310493]
68. Nawaratne R, Alahakoon D, De Silva D, O'Halloran PD, Montoye AH, Staley K, et al. Deep learning to predict energy expenditure and activity intensity in free living conditions using wrist-specific accelerometry. J Sports Sci 2021;39(6):683-690. [doi: 10.1080/02640414.2020.1841394] [Medline: 33121379]
69. Nnamoko N, Cabrera-Diego LA, Campbell D, Sanders G, Fairclough SJ, Korkontzelos I. Personalised accelerometer cut-point prediction for older adults' movement behaviours using a machine learning approach. Comput Methods Programs Biomed 2021;208:106165. [doi: 10.1016/j.cmpb.2021.106165] [Medline: 34118492]
70. O'Driscoll R, Turicchi J, Hopkins M, Duarte C, Horgan GW, Finlayson G, et al. Comparison of the validity and generalizability of machine learning algorithms for the prediction of energy expenditure: validation study. JMIR Mhealth Uhealth 2021;9(8):e23938 [FREE Full text] [doi: 10.2196/23938] [Medline: 34346890]
71. Pober DM, Staudenmayer J, Raphael C, Freedson PS. Development of novel techniques to classify physical activity mode using accelerometers. Med Sci Sports Exerc 2006;38(9):1626-1634. [doi: 10.1249/01.mss.0000227542.43669.45] [Medline: 16960524]
72. Skjødt M, Brønd JC, Tully MA, Tsai L, Koster A, Visser M, et al. Moderate and vigorous physical activity intensity cut-points for hip-, wrist-, thigh-, and lower back worn accelerometer in very old adults. Scand J Med Sci Sports 2025;35(1):e70009. [doi: 10.1111/sms.70009] [Medline: 39753998]
73. Staudenmayer J, Pober D, Crouter S, Bassett D, Freedson P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. J Appl Physiol (1985) 2009;107(4):1300-1307 [FREE Full text] [doi: 10.1152/japplphysiol.00465.2009] [Medline: 19644028]
74. Staudenmayer J, He S, Hickey A, Sasaki J, Freedson P. Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. J Appl Physiol (1985) 2015;119(4):396-403 [FREE Full text] [doi: 10.1152/japplphysiol.00026.2015] [Medline: 26112238]
75. Tsanas A. Accurately inferring physical activity levels and sleep from wrist-worn actigraphy recordings with sample rates as low as 10 hz. IEEE Access 2025;13:27257-27267. [doi: 10.1109/access.2025.3539278]
76. Walmsley R, Chan S, Smith-Byrne K, Ramakrishnan R, Woodward M, Rahimi K, et al. Reallocation of time between device-measured movement behaviours and risk of incident cardiovascular disease. Br J Sports Med 2021;56(18):1008-1017 [FREE Full text] [doi: 10.1136/bjsports-2021-104050] [Medline: 34489241]
77. Wullems JA, Verschueren SMP, Degens H, Morse CI, Onambélé GL. Performance of thigh-mounted triaxial accelerometer algorithms in objective quantification of sedentary behaviour and physical activity in older adults. PLoS One 2017;12(11):e0188215 [FREE Full text] [doi: 10.1371/journal.pone.0188215] [Medline: 29155839]
78. Wullems JA, Verschueren SMP, Degens H, Morse CI, Onambélé-Pearson GL. Concurrent validity of four activity monitors in older adults. Sensors (Basel) 2024;24(3):895 [FREE Full text] [doi: 10.3390/s24030895] [Medline: 38339613]




XSL•FO
RenderX

79. Zhou L, Qu X, Zhang T, Wu J, Yin H, Guan H, et al. Prediction of pediatric activity intensity with wearable sensors and bi-directional LSTM models. Pattern Recognit Lett 2021;152:166-171. [doi: 10.1016/j.patrec.2021.08.030]
80. Cescon M, Choudhary D, Pinsker JE, Dadlani V, Church MM, Kudva YC, et al. Activity detection and classification from wristband accelerometer data collected on people with type 1 diabetes in free-living conditions. Comput Biol Med 2021;135:104633 [FREE Full text] [doi: 10.1016/j.compbiomed.2021.104633] [Medline: 34346318]
81. Lyden K, Keadle SK, Staudenmayer J, Freedson PS. A method to estimate free-living active and sedentary behavior from an accelerometer. Med Sci Sports Exerc 2014;46(2):386-397 [FREE Full text] [doi: 10.1249/MSS.0b013e3182a42a2d] [Medline: 23860415]
82. Chan S, Hang Y, Tong C, Acquah A, Schonfeldt A, Gershuny J, et al. CAPTURE-24: A large dataset of wrist-worn activity tracker data collected in the wild for human activity recognition. Sci Data 2024;11(1):1135 [FREE Full text] [doi: 10.1038/s41597-024-03960-3] [Medline: 39414802]
83. Wang L. Recognition of human activities using continuous autoencoders with wearable sensors. Sensors (Basel) 2016;16(2):189 [FREE Full text] [doi: 10.3390/s16020189] [Medline: 26861319]
84. Ni Q, Fan Z, Zhang L, Nugent CD, Cleland I, Zhang Y, et al. Leveraging wearable sensors for human daily activity recognition with stacked denoising autoencoders. Sensors (Basel) 2020;20(18):5114 [FREE Full text] [doi: 10.3390/s20185114] [Medline: 32911780]
85. Xu D, Zhou H, Jie T, Zhou Z, Yuan Y, Jemni M, et al. Data-driven deep learning for predicting ligament fatigue failure risk mechanisms. Int J Mech Sci 2025;301:110519 [FREE Full text] [doi: 10.1016/j.ijmecsci.2025.110519]
86. Trost SG. State of the art reviews: measurement of physical activity in children and adolescents. Am J Lifestyle Med 2007;1(4):299-314. [doi: 10.1177/1559827607301686]
87. McKenzie TL, van der Mars H. Top 10 research questions related to assessing physical activity and its contexts using systematic observation. Res Q Exerc Sport 2015;86(1):13-29. [doi: 10.1080/02701367.2015.991264] [Medline: 25664670]
88. Cleland I, Kikhia B, Nugent C, Boytsov A, Hallberg J, Synnes K, et al. Optimal placement of accelerometers for the detection of everyday activities. Sensors (Basel) 2013;13(7):9183-9200 [FREE Full text] [doi: 10.3390/s130709183] [Medline: 23867744]


## Abbreviations

**AI:** artificial intelligence
**ANN:** artificial neural network
**BiLSTM:** bidirectional long short-term memory
**CNN:** convolutional neural network
**CNN-BiLSTM:** convolutional neural network and bidirectional long short-term memory
**CNN-LSTM:** convolutional neural network and long short-term memory
**DL:** deep learning
**DO:** direct observation
**DT:** decision tree
**IC:** indirect calorimetry
**LSTM:** long short-term memory network
**MET:** metabolic equivalent
**ML:** machine learning
**MPA:** moderate physical activity
**MVPA:** moderate-to-vigorous physical activity
**OSU:** Oregon State University
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**PRISMA-ScR:** The Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
**RF:** random forest
**RQ:** research questions
**SVM:** support vector machine
**UOULU:** University of Oulu
**ViT:** vision transformer
**VPA:** vigorous physical activity



XSL•FO
RenderX

*Edited by N Cahill; submitted 27.Apr.2025; peer-reviewed by S Hansun, Y Gu; comments to author 16.Sep.2025; accepted 26.Oct.2025; published 08.Jan.2026.*

*Please cite as:*
*Zi Y, van de Ven SRB, de Geus EJC, Chen P*
*Machine and Deep Learning for Detection of Moderate-to-Vigorous Physical Activity From Accelerometer Data: Systematic Scoping Review*
*Interact J Med Res 2026;15:e76601*
*URL: https://www.i-jmr.org/2026/1/e76601*
*doi:10.2196/76601*
*PMID:*

©Yahua Zi, Sjors RB van de Ven, Eco JC de Geus, Peijie Chen. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 08.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Review

# Treatment of Gender in Research on Intervention Programs Targeting Social Isolation and Loneliness Among Older Adults: Scoping Review

Kenta Nomura[1], PhD; Naoto Kiguchi[2], MA; Eisuke Inomata[3], MA; Takeshi Nakamachi[4], MA; Norikazu Kobayashi[3], PhD

[1]Graduate School of Rehabilitation, Mejiro University, Saitama, Japan

[2]Department of Occupational Therapy, Faculty of Health and Medical Sciences, Ibaraki Prefectural University of Health Sciences, Ibaraki, Japan

[3]Department of Occupational Therapy, Graduate School of Human Health Science, Tokyo Metropolitan University, Tokyo, Japan

[4]Department of Psychology and Social Welfare, Faculty of Psychology and Social Welfare, Seigakuin University, Saitama, Japan

**Corresponding Author:**
Kenta Nomura, PhD
Graduate School of Rehabilitation
Mejiro University
320 Ukiya
Iwatsuki-ku, Saitama-shi
Saitama, 339-8501
Japan
Phone: 81 487972131
Fax: 81 487972132
Email: nomura@mejiro.ac.jp

## Abstract

**Background:** Social isolation and loneliness have considerable health implications. Research indicates that older men are generally more susceptible to social isolation compared with women, highlighting the need to integrate gender-responsive approaches in the development and implementation of interventions for mitigating social isolation and loneliness in later life.

**Objective:** This study aimed to conduct a review of intervention programs targeting social isolation and loneliness, focusing on gender-specific considerations. Specifically, it aims to examine the gender composition (male-to-female ratio) of participants in intervention programs and identify and analyze intervention strategies that demonstrate gender-sensitive effectiveness.

**Methods:** A scoping review was conducted as per the Joanna Briggs Institute manual for evidence synthesis. A comprehensive literature search, including hand searching, was conducted across 6 English-language databases, PubMed, MEDLINE, Cochrane, CINAHL, ScienceDirect, and Web of Science, for papers and reports published in 2013-2023. The authors, country, subjects, research design, intervention method, results, and mentions of gender for each included document were presented.

**Results:** The study identified 1282 papers and reports, of which 10 were selected for analysis. Only 1 study reported a higher number of male participants compared with female ones; in contrast, all other studies included predominantly female samples. The studies assessed outcomes based on 2 indicators of social isolation, 4 indicators of loneliness, and 29 other indicators. Exercise and workshops proved effective for social isolation and loneliness, while meditation and laughter therapy were effective for loneliness. The intervention with the highest percentage of male participants (264/323, 82%) was a customized meditation program. Conversely, physical activities, social support, and community-based group health classes drew more female participants. In total, 8 studies did not mention gender in the discussion section, and none considered gender-specific issues in formulating research objectives and outcomes.

**Conclusions:** Research on social isolation and loneliness has generally ignored the influence of gender. The review also indicated a gender bias in participant selection, with women markedly overrepresented in study samples. The study found that women tend to prefer interventions emphasizing conversations, shared experiences, and emotional exchange. In contrast, men showed the highest participation in a meditation program focused on self-dialogue, which required minimal interaction. Importantly, interventions aimed at promoting social interaction or participation are unlikely to succeed without consideration of gender-specific issues. Therefore, systematically identifying conditions necessary for effective interventions that target older men is crucial for guiding future research and program development.



XSL•FO
RenderX

**Trial Registration:** Open Science Framework 10.17605/OSF.IO/83JQF; hhttps://osf.io/83jqf/overview

***(Interact J Med Res 2026;15:e72281)*** doi:10.2196/72281

**KEYWORDS**

social isolation; loneliness; intervention; program; gender; men; women; older; review; remote; in-person; scoping review

# Introduction

## Background

Social isolation is defined as "the objective lack or paucity of social contacts and interactions with family members, friends, or the wider community" [1]. Factors associated with social isolation include advanced age, male gender, presence of depressive symptoms, and low socioeconomic status [2]. The concept of social isolation is closely related to loneliness. According to Valtorta and Hanratty [1], loneliness is "a subjective negative feeling associated with a perceived lack of a wider social network (social loneliness) or the absence of a specific desired companion (emotional loneliness)." Scholars have defined social isolation as a lack of interaction with others and distinguished loneliness from subjective loneliness [3]. Although social isolation and loneliness are considered closely related, their definitions differ. Some studies have found only a weak correlation [4,5]; socially isolated individuals are not necessarily lonely, and vice versa. Despite these distinct definitions and realities, social isolation and loneliness are often studied together as overlapping concepts. Newall [6] also proposed that they should be examined in conjunction.

A growing body of evidence indicates that gender is a crucial factor influencing older adults' experiences of social isolation and loneliness. For example, older men (ie, those aged 65 years and older) in Japan are more likely to experience social isolation compared with their female counterparts [7,8]. In a cross-sectional study, Nomura and Kobayashi [9] discovered that gender can influence the isolation-prevention strategies used by community-dwelling older people in Japan. The authors found that satisfaction with social activities (ie, the degree of satisfaction with activities involving participation in groups and organizations and interpersonal activities with others) and strategies for interactions helped older Japanese women maintain relationships with others. However, for older Japanese men, only satisfaction with social activities contributed to maintaining social relationships. Some studies also identify the need to focus intervention efforts on men [10]. Men are more likely to feel lonely compared with women for one reason: when confronted with challenges or distressing situations, women tend to respond emotionally and engage in conversation with others, whereas men tend to avoid stressors and neglect talking to others about them [11]. These findings underscore the importance of considering gender in the development of interventions targeting social isolation. Nevertheless, few intervention programs have appropriately accounted for gender. Furthermore, a comprehensive overview of findings on the salience of gender in social isolation research is lacking. Moreover, the treatment of gender in research on social isolation is inconsistent, given that researchers adopt different approaches to this issue.

Barreto et al [12] argued that research on loneliness has failed to address gender beyond the binary categories of male and female, thereby obscuring important influences. Other meta-analyses have also indicated a correlation between loneliness and gender. For example, Pollak et al [13] identified gender as a predictor and risk factor for loneliness and functional decline, while Haihambo et al [14] suggested that gender differences exist in both loneliness and social adaptability. However, these studies represent only one dimension of the analysis and fail to specifically focus on gender. Although gender is a critical determinant in understanding social isolation and loneliness, the prevalence of gender-focused interventions remains underexplored.

Several studies have examined intervention programs designed to address social isolation and loneliness. For example, Milligan [15] conducted a scoping review of Men's Sheds and gender interventions to assess their impact on the health and well-being of older men. Men's Sheds are community-based, mutual self-help groups established primarily in Australia and the United Kingdom, among other countries. These initiatives were intended to alleviate social isolation and loneliness among retired men by fostering purpose, companionship, and shared activities [16]. The review identified limited evidence supporting the impact of Men's Sheds and other gender-focused social activities on the mental health and well-being of older men. However, this paper was published in March 2015; therefore, it does not cover research published after 2014. Fakoya et al [17] conducted a scoping review of studies on loneliness and social isolation interventions targeting older adults to identify effective interventions. The sample comprised 33 papers published until 2018. Findings indicated that the individual nature of experiences of social isolation and loneliness may impede the development and implementation of standardized interventions. The study concluded that no universal approach exists for addressing social isolation and loneliness, and interventions must be tailored to specific individuals and groups, as well as the severity of loneliness. Although existing studies encompass diverse interventions, a key limitation is that none explicitly considered gender as an influencing variable. Although a few studies address the potential impact of gender on interventions targeting the alleviation of social isolation and loneliness, the actual prevalence of such gender-focused interventions remains unclear.

## Objective

This study aims to comprehensively map intervention programs targeting social isolation and loneliness among older adults, with a particular emphasis on the integration of gender considerations in these programs. Specifically, it intends to (1) identify the male-to-female ratio of participants across studies and (2) determine effective intervention strategies that are responsive to gender differences. The results are expected to



XSL•FO
RenderX

facilitate the development and implementation of new gender-focused intervention programs.

## Methods

### Research Design

A scoping review was conducted in accordance with the Joanna Briggs Institute (JBI) manual for evidence synthesis (hereafter, JBI framework) [18]. The JBI framework defines the purpose of a scoping review as identifying and analyzing knowledge gaps and examining how research is conducted within a given field [18]. The decision to conduct a scoping review was aimed at mapping the existing evidence and identifying areas that require further exploration rather than synthesizing findings, as customary in a systematic review. The review protocol was registered with the Center for Open Science Framework. No deviations were observed between the registered protocol and the content presented in this study. The inclusion criteria were based on the following definitions of participants, concept, context, and types of sources of evidence.

### Participants

The participants were older men and women residing in a community, regardless of whether they reported being socially isolated. The review also focused on older adults but did not strictly define the participants' ages. Thus, age was excluded as an eligibility criterion given the diverse definitions of "elderly" across countries and studies.

### Concept

This study focused on the implementation of strategies for addressing social isolation and loneliness. However, studies examining the correlation of social isolation and loneliness with curfew restrictions during the COVID-19 pandemic were excluded. This study focused on cases of social isolation and loneliness in which underlying causes remained unclear and included gender as a variable. The present-day concept of gender is multifaceted; nevertheless, it is often reduced to a binary model that assigns specific biological and behavioral traits to men and women.

### Context

Social isolation and loneliness are influenced by various social factors, including geographic location, religious affiliation, and racial identity. This study does not impose any specific limitations regarding such factors; instead, it encompasses diverse community-based interventions.

### Types of Sources

This review covered various intervention studies, such as randomized controlled trials, nonrandomized controlled trials, and before-and-after studies. Quantifying the effectiveness of interventions targeting social isolation and loneliness is challenging. Therefore, the review also included websites, reports, and research papers while excluding observational studies, qualitative studies, reviews, and conference abstracts. The rationale for incorporating websites was twofold. First, a significant number of interventions that target social isolation and loneliness have not been disseminated through academic journals. Second, existing research on social isolation and loneliness has predominantly focused on academic publications, with nonacademic sources receiving minimal representation.

### Literature Search Strategy

The literature search strategy in this study was developed in consultation with a librarian at Mejiro University. A search was conducted on 6 databases (PubMed, MEDLINE, Cochrane, CINAHL, ScienceDirect, and Web of Science) on May 30, 2023, to identify relevant papers published within an 11-year period (2013-2023). The following terms were searched in combination: intervention, social isolation, loneliness, elderly, older adults, older people, older peoples, sex characteristics, sex, gender, and difference (Textbox 1). In addition to database searches, hand searches were conducted on the websites of 3 journals that frequently publish research on social isolation and loneliness: *Health and Social Care in the Community*, *Journal of Aging and Health*, and *Ageing & Society*. Furthermore, a hand search for gray literature was conducted on 2 websites, the Social Care Institute for Excellence and Connect2Affect, for the terms "social isolation" or "loneliness," with the publication period limited to 2013-2023 (Multimedia Appendix 1).

**Textbox 1.** Search strategy.

"(intervention)

and (((((social isolation)or(loneliness))

and (((aged)or(elderly))or((""older people"")or(""older peoples""))))

and((sex characteristics)or(((sex)or(gender))and(difference)))))"

### Literature Selection

Data extraction was performed by KN. All identified academic studies and relevant texts were uploaded to Rayyan [19]. Duplicate papers were omitted, followed by a screening. KN and EI conducted the initial and full-text screening of the literature, while NK was consulted in cases of disagreement. The sample included English-language papers featuring intervention programs for social isolation and loneliness. The following texts were excluded: conference abstracts; studies without mention of social isolation or loneliness in their objectives, methods, results, or discussion; and studies that focused on COVID-19 countermeasures.

### Data Analysis

The selected studies were presented in the form of a table that included the following information for each study: author, country of origin, participants, intervention method, outcome, and gender. An additional table was constructed to illustrate the male-to-female participant ratio per study and to classify the



XSL•FO
RenderX

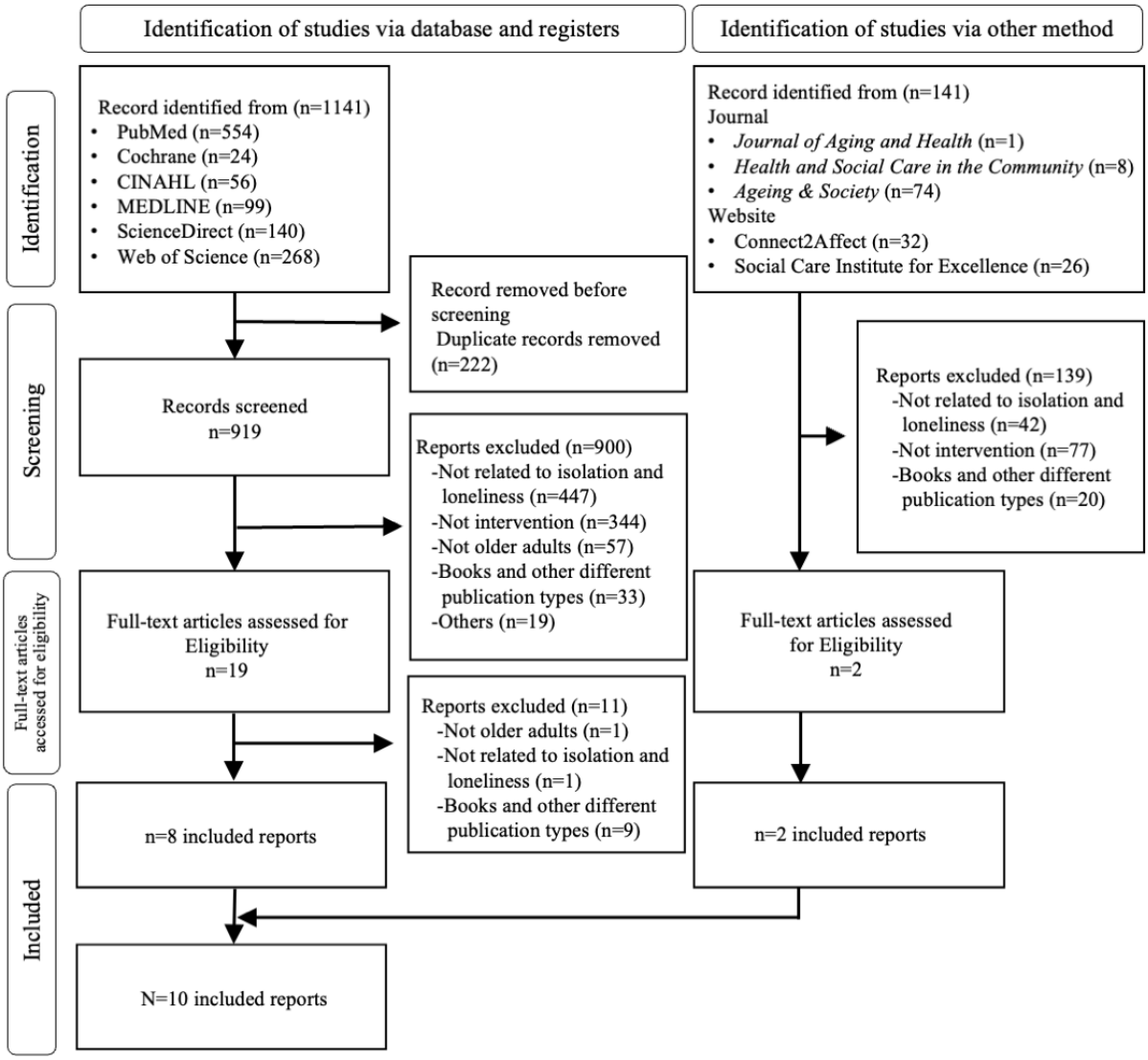
type of intervention (remote or in-person, group-based or individual, or a combination of these formats). Furthermore, the measures and outcome indicators used to evaluate the interventions were presented in tabular form. To enhance clarity and comparability, the primary outcome concepts were systematically categorized and organized according to their thematic relevance.

## Results

### Characteristics of the Selected Studies

The literature selection process was visualized as a flowchart in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist (Figure 1; Multimedia Appendix 2) [20]. The database search yielded 1141 studies, of which 222 were duplicates and were therefore excluded. The screening process yielded 19 studies, 8 [21-28] of which were included in the analysis. Of the 900 studies that were excluded during the screening process, 447 were unrelated to social isolation and loneliness, while 344 were not intervention studies. The hand search yielded 141 studies, of which 2 were included (Table 1). A total of 10 studies [21-30] were included in the final analysis. These studies described detailed interventions conducted in the United States (n=3 [23,27,28]), the United Kingdom (n=2 [29,30]), Singapore (n=1 [26]), Turkey (n=1 [25]), Spain (n=1 [21]), Ireland (n=1 [23]), and India (n=1 [22]; Table 1).

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of study selection.






XSL•FO
RenderX

**Table 1.** Summary of included studies.

| First author, date, and country | Participants (final analyzed individuals) | Intervention description | Result | Gender references |
|---|---|---|---|---|
| Ngiam et al, 2022 [26], Singapore | Older adults who resided in the southeast region of Singapore, aged >55 years, belonging to a lower socioeconomic status. 138 were included for analysis (59 male and 79 female). | Project Wire Up was a volunteer-led, one-on-one, goal-directed, and home-based digital literacy program: (1) equipped with smartphones and internet connection; (2) trained by volunteers for 6 sessions (1 to 2 hours per session) over 3 months that were held in the older adults' homes; and (3) digitally connected to existing social networks. | There were significant improvements in digital literacy scores in the intervention group as compared to controls (mean difference 2.28, 95% CI 1.37-3.20; *P*<.001). There was no statistically significant difference in the University of California, Los Angeles 3-item loneliness scale, Lubben Social Network Scale-6, Personal Wellbeing Score, or EQ-5D Utility and visual analog scale score. | No mention |
| Santos-Olmo et al, 2022 [21], Spain | The sample (N=68: 18 male, 50 female) was 65 years of age or older; lived alone or lived with other people older than 65 years; had uncovered social and/or health needs; have little or no social support network; refused the assistance offered to cover their needs from the normalized social and/or health services. | The Psychological Support Service for Socially Isolated Elderly People (PSIE) facilitates contact between the elderly person and social and health services (both primary and specialized care) in their area, so that they can receive the care they need on each occasion. | Concerning the total score of the Health and Psychosocial Functioning (Total HoNOS65+, *t*=10.12, *P*<.001, Cohen *d*=1.49) and the Spanish adaptation of the Camberwell Needs Assessment Questionnaire for the Elderly (CANE; average number of unmet needs, *t*=19.99, *P*<.001, Cohen *d*=2.31), significant differences were also observed between pre- and posttreatment. Statistically significant changes were obtained in Global Assessment of Functioning (GAF; *t*=4.06, *P*<.001, Cohen *d*= 0.6) and the Spanish adaptation of the WHO Disability Assessment Short Scale (WHO-DAS-S; A. personal care and survival, *t*=8.82, *P*<.001, Cohen *d*=1.3). | The gender variable does not seem to have an influence on any of the outcome measures studied, except for alcohol use. PSIE has achieved positive results in both men and women, without gender appearing to be a relevant variable in these results. |
| Dodge et al, 2022 [27], United States | Participants with normal cognition or mild cognitive impairment were recruited from Portland, OR, and Detroit, MI. Key inclusion criteria included (1) age 75 years or older, (2) socially isolated. One hundred eighty-six participants (86 with normal cognition and 100 (52.8%) with mild cognitive impairment) were randomized into the experimental (n=94) or the control group (n=92). | The intervention group had video chats with trained study staff for 30 minutes per day, 4 times per week for 6 months (high dose), and then twice per week for an additional 6 months (maintenance dose). Both intervention and control groups received a phone call once per week (approximately 10 minutes duration) to assess changes in health and social activities. | After the induction period, the experimental group had higher global cognitive test scores (Montreal Cognitive Assessment [primary outcome]; 1.75 points [*P*=.03]). After induction, participants in the experimental group with normal cognition had higher language-based executive function (semantic fluency test [secondary outcome]; 2.56 points [*P*=.03]). At the end of the maintenance period, participants in the experimental group with mild cognitive impairment had higher encoding function (Craft Story immediate recall test [secondary outcome]; 2.19 points [*P*=.04]). | No mention |
| Pandya et al, 2021 [22], India | Intervention group older adults ($IN_2$=166, 136 males, 30 females), the control group ($CN_2$=157, 128 males, 29 females), who underwent no intervention | The key features of the meditation program were (1) postures interspersed with relaxation, (2) slowness in movements, and (3) inner watchful awareness. | There were significant mean differences in the posttest scores on loneliness, well-being, life satisfaction, and contentment outcomes of the intervention group, with high observed effect sizes (Cohen *d* range=2.43-8.78; *P*≤.01). The intervention group older adults reported that they were less lonely and experienced greater well-being, life satisfaction, and contentment posttest ($\eta p^2$ =0.71-0.78; *P*≤.01). | Men, middle class, married, and cohabitating participants, who also comprised a majority of the sample, were less lonely at the pretest phase as compared to women, upper class, single, and living alone. Gender was an important factor, and results showed that retired South Asian men were less lonely and more satisfied pretest as well as posttest. |



XSL•FO
RenderX

| First author, date, and country | Participants (final analyzed individuals) | Intervention description | Result | Gender references |
|---|---|---|---|---|
| Mays et al, 2021 [23], United States | Participants (n=382, 63 male, 315 female) were offered enrollment based on the following inclusion criteria: aged 50 years or older, community-dwelling, able to complete questionnaires, able to consent to participate in the study, and able to communicate in English. | Participants met with the program coordinator and selected from 1 of the 4 evidence-based programs: Tai Chi for Arthritis, Enhance Fitness, the Arthritis Foundation Exercise Program, and the Healthier Living Workshop. Group health class lasted 6 to 8 weeks. | Older adults who met with a health coach and participated in a single session of community health programs reported decreased loneliness (ER[a] 0.931, 95% CI 0.895-0.968; *P*<.001) and social isolation (ER 1.033, 95% CI 1.016-1.050; *P*<.001) at 6 months post participation, compared to their baseline scores. | No mention |
| Zamir et al, 2020 [29], United Kingdom | Twenty-two residents aged ≥65 years (5 male, 17 female) across 3 British care homes | Twenty-two residents engaged with each other using "Skype quiz" sessions with the support of staff once a month over an 8-month trial. Residents met other residents from the 3 care homes to build new friendships and participate in a 30-minute quiz session facilitated by 8 staff members. | Analysis of the field notes revealed 5 themes of: residents with dementia remember faces, not technology, inter- and intraconnectedness, regaining sense of self and purpose, situational loneliness overcome, and organizational issues create barriers to long-term implementation. | No mention |
| Lawlor et al, 2019 [24], Ireland | Participants (n=40) from older (aged≥50 years) women's groups from 4 different community centers | Intervention consisted of 3 face-to-face group education sessions, encouragement to enlist the support of a buddy (eg, spouse, partner, friend, or a group member), an information pack, and the option of weekly telephone contact. Each education session lasted approximately 20 minutes. | 87% (40/46) of women consented to participate, and 78% (31/40) attended all education sessions. Few participants provided valid accelerometer data, but 63% (25/40) completed the HADS[b] questionnaire at all time points. 85% of participants (34/40) were somewhat or very satisfied with their involvement in the study. | No mention |
| Alıcı et al, 2018 [25], Turkey | The study participants were older adults living in 2 nursing homes set up by foundations located in the capital of Turkey. Their ages were more than 65 years. A total of 50 older adults formed the intervention group (n=20; 9 males, 11 females) and control group (n=30; 14 males, 16 females). | Laughter therapy was conducted by the principal investigator 2 days a week, with one application during each session. The program involved performing yoga, breathing, and physical exercises, as well as laughter therapy. The program continued for 5 weeks for a total of 10 applications. The control group received no intervention. | A statistically significant difference (*P*<.001) between mean De Jong Gierveld Scale scores of the intervention (mean 7.15, SD 1.755) and control groups (mean 15.63, SD 5.027) was observed after the intervention. Median De Jong Gierveld Scale scores were significantly lower in the intervention group than in the control group. After therapy, the social loneliness score was significantly lower in the intervention group (mean 3.10, SD 1.553; *P*<.001) than in the control group (mean 6.90, SD 3.100). Post therapy, the emotional loneliness score was significantly lower (*P*<.001) in the intervention group (mean 4.05, SD 1.538) than in the control group (mean 8.73, SD 2.599). | No mention |
| Dodge et al, 2015 [28], United States | Eighty-three individuals participated (41 in the intervention group and 42 in the control group). Participants aged 70 years or older were included. | Daily 30-minute face-to-face communications were conducted during a 6-week trial period in the intervention group. The control group received only a weekly telephone interview. | Among the cognitively intact participants, the intervention group improved more than the control group on a semantic fluency test (*P*=.003) at the posttrial assessment and a phonemic fluency test (*P*=.004) at the 18-week assessments. Among those with mild cognitive impairment, a trend (*P*=.04) toward improved psychomotor speed was observed in the intervention group. No difference was found between intervention and control groups in the pre- to posttrial changes in the loneliness score, the secondary outcome. | No mention |



XSL•FO
RenderX

| First author, date, and country | Participants (final analyzed individuals) | Intervention description | Result | Gender references |
|---|---|---|---|---|
| Hind et al, 2014 [30], United Kingdom | The eligibility criteria included being aged ≥ 75 years, living independently, and having reasonable cognition. Fifty-six participants (control n=30, intervention n=26) were included in the intention-to-treat analysis group. | Manualized the telephone friendship (TF) with standardized training: (1) one-to-one befriending: 10- to 20-minute calls once per week for up to 6 weeks made by a volunteer befriender, followed by (2) TF groups of 6 participants: 1-hour teleconferences once per week for 12 weeks facilitated by the same volunteer. Control: usual health and social care provision. | The 2 groups were reasonably well matched with respect to baseline demographic characteristics. At 6 months post randomization, the SF-36 mental health mean (SD) scores were 77.5 (18.4) in the intervention group and 70.7 (21.2) in the control group, with a mean difference of 6.5 (95% CI –3.0 to 16.0); after adjusting for age, sex, and baseline scores, the mean difference was 9.5 (95% CI 4.5-14.5). In summary, over the 6-month follow-up period, there was no change in the SF-36 mental health scores in the intervention group, but there was a decline or deterioration in scores in the control group. | No mention |

[a]ER: estimated ratio.

[b]HADS: Hospital Anxiety and Depression Scale.

The literature search was conducted using 2 complementary approaches: database and hand searches. Database searches were performed across the 6 abovementioned databases, yielding 1141 records, of which 8 studies [21-28] met the inclusion criteria. Hand searches were conducted across 3 journals and 2 websites, identifying 141 records, of which 2 [29,30] studies were eligible for inclusion.

In total, 10 studies [21-30] were included in the final analysis.

Table 1 presents a summary of the 10 papers included in the final analysis. All studies were published between 2013 and 2023. The United States and the United Kingdom were the countries with the highest representation in terms of intervention programs. Participants were generally older adults aged ≥50 years, with the majority being ≥65 years and predominantly female.

## Characteristics of Participants

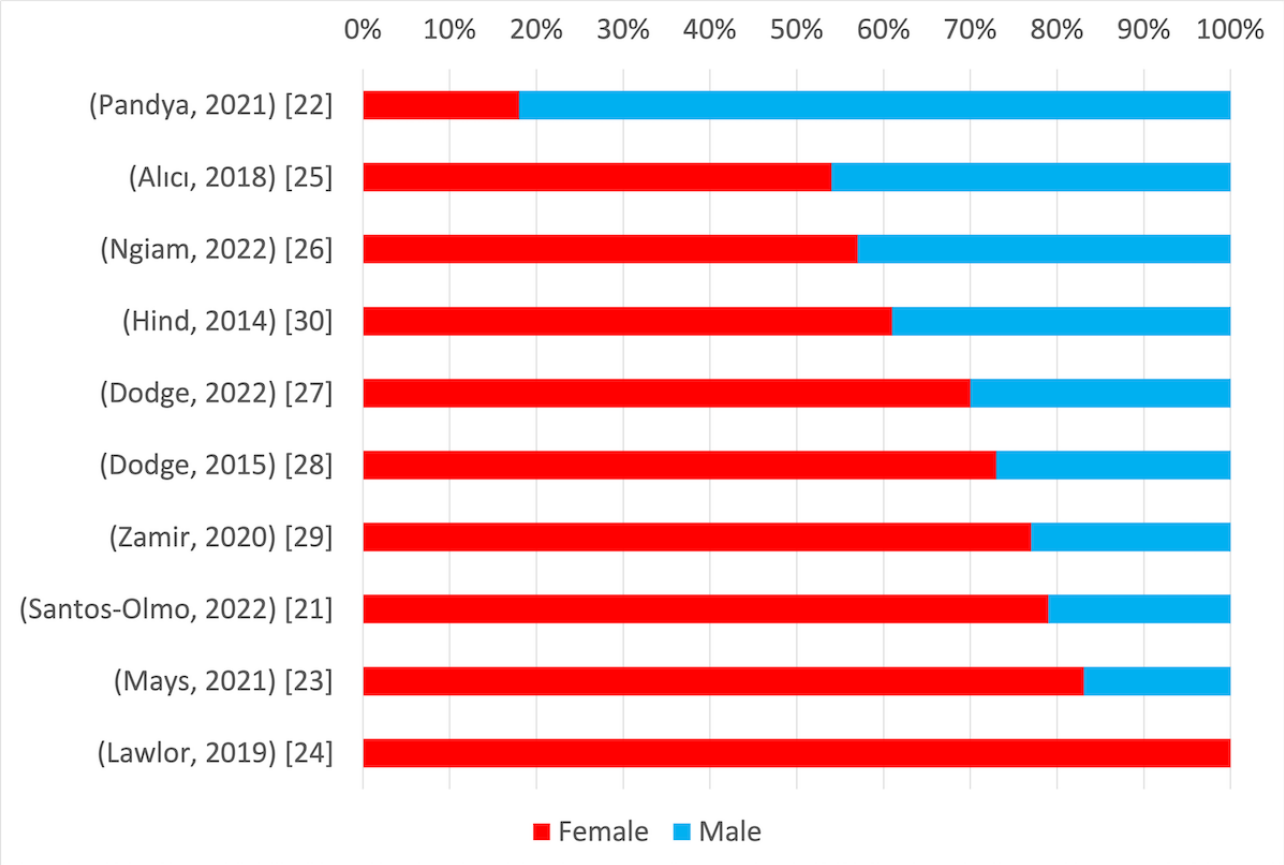
The 10 included studies [21-30] involved 1343 older adult participants, with sample sizes ranging from 22 to 382 (mean 134.3, SD 119.2; median 73, IQR 51.5-174). The proportion of female participants varied per study: approximately 10%-20%, 50%-60%, 60%-70%, 70%-80%, 80%-90%, and 100% women participated in 1 [26], 2 [21,27], 1 [22], 4 [23-25,29], 1 [28], and 1 [30] study, respectively (Figure 2 [21-30]). Overall, 8 documents incorporated the term “social isolation” or “loneliness” in their objectives, methods, or results. However, these documents did not explicitly address gender in the discussion section. In contrast, a few documents incorporated gender-related content in the discussion section. Studies referencing gender examined its impact on intervention effectiveness [21] and loneliness levels [22]. Among the included studies, the most common intervention method for those with the highest proportion of male participants (82%) was a personalized meditation program [22]. Conversely, studies involving women used different methods, including community-based group health classes [23] and physical activity and social support [24].

Only 1 [22] study reported a higher number of male participants compared with female ones; the remaining studies had a majority of female participants. One study [24] exclusively targeted older women, whereas no study focused solely on older men.



XSL•FO
RenderX

**Figure 2.** Gender ratio of participants [21-24, 26-30].




## Intervention Methods and Means

The studies used 4 types of study designs: randomized controlled trials (n=5), nonrandomized control trials (n=1), pre-post outcome studies (n=3), and action research (n=1). They assessed 5 types of in-group programs: a program intended to enhance access to social and health services [21], a meditation program [22], an exercise and workshop [23], a physical activity program [24], and a laughter therapy program [25]. Overall, 3 studies discussed individualized remote programs, including a digital literacy program [25], a video chatting program [27], and a face-to-face communication program [28]. Another study assessed a remote group program that used a Skype (Microsoft) quiz format [29]. Furthermore, 1 study evaluated a remote individual program and a remote group program based on telephone friendship [30] (Table 2). The exercise workshop [23] effectively alleviated social isolation and loneliness indicators; meanwhile, the meditation [22] and laughter therapy programs [25] effectively mitigated loneliness (Table 1).

The interventions were nearly evenly divided between “remote” (n=5) and “in-person” (n=5) formats. Notably, no individual in-person programs were reported; instead, group-based in-person interventions were the most common. In-person programs typically involved strategies such as facilitating access to social resources, meditation, exercise, and workshops. Remote programs included activities such as digital literacy training, video-based communication, and online conversational sessions.



XSL•FO
RenderX

**Table 2.** Intervention methods.

| Remote | Face-to-face |
|---|---|
| **Individual** | |
| • Digital literacy program (Ngiam et al [26], 2022)<br>• Video chats (Dodge et al [27], 2022)<br>• Face-to-face communications (Dodge et al [28], 2015) | —[a] |
| **Group** | |
| • Skype quiz (Zamir et al [29], 2020) | • Improving access to social and health services (Santos-Olmo et al [21], 2022)<br>• Meditation program (Pandya [22], 2021)<br>• Physical activity (Lawlor et al [24], 2019)<br>• Laughter therapy (Alıcı et al [25], 2018) |
| **Individual and group** | |
| • Telephone friendship (Hind et al [30], 2014) | • Exercise and workshops (Mays et al [23], 2021) |

[a]Not available.

## Measurement of Effectiveness

The 10 studies [21-30] assessed outcomes based on 2 indicators of social isolation, 4 indicators of loneliness, and 29 other indicators (Table 3). Among these, outcomes with positive effects were observed for the Duke Social Support Index [23] for social isolation, the 3-item loneliness scale of the University of California, Los Angeles [23], the De Jong Gierveld Loneliness Scale [25], and the 6-item De Jong Gierveld Loneliness Scale [22].

The selected studies presented relatively few measures for assessing social isolation and loneliness; however, they used a broad range of indicators to evaluate cognitive function, well-being, satisfaction, and other psychosocial domains. For outcomes related to social isolation or loneliness and cognitive function, the number of studies that reported significant effects was approximately equal to those reporting no effects.



XSL•FO
RenderX

**Table 3.** Outcomes and effectiveness.

| Categories and outcomes | Effective | Ineffective |
|---|---|---|
| **Social isolation** | | |
| LSNS-6[a] | • —[b] | Ngiam et al [26], 2022 |
| DSSI[c] | • Mays et al [23], 2021 | — |
| **Loneliness** | | |
| UCLA-3[d] | • Mays et al [23], 2021 | Ngiam et al [26], 2022 |
| The De Jong Gierveld Loneliness Scale score | • Alıcı et al [25], 2018 | — |
| 3-item loneliness scale | — | Dodge et al [27], 2015 |
| DJGLS-6[e] | • Pandya et al [22], 2021 | — |
| **Other** | | |
| **Cognition** | | |
| Trail making test | — | Dodge et al [28], 2015 |
| Mini mental state examination | — | Dodge et al [28], 2015 |
| Cogstate computerized tests | — | Dodge et al [28], 2015 |
| The Stroop test | — | Dodge et al [28], 2015 |
| MoCA[f] | • Dodge et al [27], 2022 | — |
| Craft Story immediate recall test | • Dodge et al [27], 2022 | — |
| The Consortium to Establish a Registry for Alzheimer Disease word list learning | — | Dodge et al [28], 2015 |
| The Consortium to Establish a Registry for Alzheimer Disease word list delayed recall | — | Dodge et al [28], 2015 |
| CAMCI[g] | — | Dodge et al [28], 2015 |
| **Fluency** | | |
| Semantic fluency test | • Dodge et al [27], 2022)<br>• Lawlor et al [24], 2019) | — |
| The composite of verbal fluency for letters (F, A, and S) | • Dodge et al [28], 2015) | — |
| Verbal fluency for the category animals | • Dodge et al [28], 2015) | — |
| **Qualitative method** | | |
| WHO-DAS-S[h] | • Santos-Olmo et al [21], 2022 | — |
| Thematic analysis | • Zamir et al [29], 2020 | — |
| framework analysis | • Lawlor et al [24], 2019) | — |
| **Well-being** | | |
| PWS[i] | — | Ngiam et al [26], 2022 |
| WEMWBS[j] | • Pandya et al [22], 2021 | — |
| **QoL[k]** | | |
| EQ-5D | — | Ngiam et al [26], 2022 |
| SF-36[l] | • Hind et al [30], 2014 | — |
| **Psychosocial functioning** | | |



XSL•FO
RenderX

| Categories and outcomes | Effective | Ineffective |
|---|---|---|
| HoNOS65+[m] | • Santos-Olmo et al [21], 2022) | — |
| GAF[n] | • Santos-Olmo et al [21], 2022) | — |
| **Satisfaction** | | |
| SWLS[o] | • Pandya et al [22], 2021 | — |
| CLAS[p] | • Pandya et al [22], 2021 | — |
| **Emotion** | | |
| TDAS[q] | • Alıcı et al [25], 2018 | — |
| NIHTB-EB[r] | • Dodge et al [27], 2022 | — |
| **Digital literacy** | | |
| Digital literacy score | • Ngiam et al [26], 2022 | — |
| **Needs** | | |
| CANE[s] | • Santos-Olmo et al [21], 2022 | — |
| **Depression** | | |
| Hospital Anxiety and Depression Scale | — | Lawlor et al [24], 2019 |
| **fMRI[t]** | | |
| fMRI analyses | • Dodge et al [27], 2022 | — |

[a]LSNS-6: Lubben Social Network Scale-6.
[b]Not available.
[c]DSSI: The Duke Social Support Index.
[d]UCLA-3: University of California, Los Angeles 3-item loneliness scale.
[e]DJGLS-6: The De Jong Gierveld Loneliness Scale (six-items).
[f]MoCA: Montreal Cognitive Assessment.
[g]CAMCI: Computer assessment of mild cognitive impairment.
[h]WHO-DAS-S: World Health Organization Brief Disability Assessment Scale.
[i]PWS: Personal Well-being Score.
[j]WEMWBS: Warwick Edinburgh Mental Well-being Scale.
[k]QoL: quality of life.
[l]SF-36: 36-Item Short Form Health Survey.
[m]HoNOS65+: Home Health Outcome Scales for People Over 65.
[n]GAF: Global Assessment of Functioning.
[o]SWLS: The Satisfaction with Life Scale.
[p]CLAS: The Contentment with Life Assessment Scale.
[q]TDAS: The Turkish Death Anxiety Scale.
[r]NIHTB-EB: National Institutes of Health-toolbox emotional battery.
[s]CANE: Camberwell Needs Assessment Questionnaire for the Elderly.
[t]fMRI: functional magnetic resonance imaging.

# Discussion

## Principal Findings

In this review, a total of 1282 papers were identified, of which only 10 [21-30] were eligible for analysis. Studies were excluded primarily because they did not specifically examine social isolation or loneliness or were merely cross-sectional investigations that emphasized the need for interventions. Of the included studies, East Asia was represented only by Singapore; notably, no research was included from Japan despite its status as a super-aged society.

Regarding the interventions, the methods were evenly distributed between remote and in-person methods. In total, 5 [26-30] of the 10 (50%) studies [21-30] used remote interventions (eg, digital literacy and video chats), while the other 5 studies [21-25] used in-person interventions (eg, exercise, meditation, and



XSL•FO
RenderX

laughter therapy). However, none of them combined remote and in-person approaches.

A significant gender bias was observed; in 9 studies [21-24,26-30], most participants were women. Among studies in which women accounted for ≥80% or more of the participants, interventions frequently emphasized communication, shared emotional expression, and group-based activities. Conversely, the study with the highest proportion of male participants (82%) featured a meditation program characterized by self-dialogue and individual reflection.

Importantly, only 2 [21,22] of the 10 studies [21-30] discussed the impact of gender on intervention outcomes, whereas no study focused on gender-specific issues as a primary research objective. Furthermore, the use of specific indicators for social isolation and loneliness was limited; indicators were typically substituted by measures of cognitive function or quality of life, depending on the study's specific objectives.

## Comparison With Prior Work

The finding that 50% of the studies used remote interventions represents a significant increase compared with the result of a previous scoping review conducted by Fakoya et al [17]. The authors found that only 15% (5 of 33) of interventions cited in papers published between 1984 and 2018 were remote. This trend aligns with the growing efficacy and development of remote tools [31,32], likely accelerated by the global COVID-19 pandemic. However, while Sen et al [32] recommended the use of online and offline resources to enhance efficacy, this review found no studies adopting such a hybrid approach.

The predominance of female participants is consistent with previous literature emphasizing that research on social isolation and loneliness is biased toward women [33,34]. This finding contradicts evidence suggesting that older men are more prone to social isolation compared with women [7,8].

Regarding intervention preferences, the findings support existing theories on gender differences. Female participants' preference for exchanging experiences in supportive settings mirrors the findings of Venter et al [35]. In contrast, the male preference for the meditation program aligns with research proposing that men gravitate toward task-oriented activities, autonomy, and less intimate social networks, as seeking help or emotional disclosure can be perceived as vulnerability [36].

The lack of gender-specific analysis in the included studies is in contrast with the recommendations of Santos-Olmo et al [21] and Wen et al [10], who argued that future research needs to prioritize gender as a variable for predicting the effectiveness of interventions. The absence of studies conducted in Japan is also a concern, given the cultural specificity of isolation [13,37], thus highlighting a gap between Japan's urgent policy needs [38] and the available evidence base.

## Strengths and Limitations

### *Strengths*

The primary strength of this study is its explicit focus on gender. To the best of our knowledge, this is the first literature review on social isolation and loneliness that considers gender as a central variable. Additionally, relevant studies were identified using a rigorous methodology based on multidisciplinary databases.

### *Limitations*

First, the review may not have captured all relevant studies, partially because the study period overlapped with the COVID-19 pandemic, potentially causing delays in publication. Second, the literature search concluded in May 2023, excluding subsequent studies. Third, despite manual searches, some high-quality sources or community-based practices that indirectly address isolation (eg, neighborhood association activities) may have been missed. Finally, as a scoping review, no quality assessment of the included papers was conducted; thus, the quality of evidence may vary.

## Future Directions

Gender-specific interventions: a clear need exists to develop and implement new gender-focused intervention programs.

For older women: interventions should prioritize group-based activities that foster dialogue, mutual understanding, and emotional connection.

For older men, future efforts must rigorously identify criteria for interventions targeting men. Programs should incorporate activities that enable self-expression without overt emotional disclosure, such as those emphasizing skill use, personal autonomy, and a sense of contribution.

Research design: future studies must move beyond merely reporting gender demographics. Studies should include gender as the primary variable in order to analyze its impact on intervention outcomes. A reexamination of previous studies that failed to show significant effects could reveal whether the lack of gender consideration was a contributing factor.

Hybrid approaches: given the increased popularity of remote interventions and the established value of in-person interaction, future interventions should aim to combine remote and in-person methods. Leveraging the experience of online-based interventions to create hybrid models could enhance the precision and efficacy of countermeasures against social isolation and loneliness.

## Conclusions

This scoping review, which aimed to examine the composition of participants by gender and gender-sensitive strategies, confirms that research on social isolation generally underestimates gender. This finding reveals a bias in which women are markedly overrepresented in study samples, contradicting evidence that older men are more susceptible to isolation. Critically, no study considered gender-specific issues in its objectives, undermining the efficacy of general interventions. We found gender-based differences in preferences: women favored conversation and emotional exchange, while men exhibited the highest proportion of participation in self-dialogue meditation programs. Therefore, a clear need exists to develop new gender-focused intervention programs and systematically identify conditions for effective interventions that specifically target older men.



XSL•FO
RenderX

## Acknowledgments

The authors extended their gratitude to the librarians at Ibaraki Prefectural University of Health Sciences for their invaluable assistance in the design of the search strategy. No text generated by artificial intelligence was used in the writing of this study.

## Funding

This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grants-in-Aid for Scientific Research (grant JP23K12637).

## Data Availability

All data generated or analyzed during this study are included in this published paper.

## Authors' Contributions

Conceptualization: KN (lead) and N Kiguchi (supporting)
Data curation: KN
Formal analysis: KN (lead) and EI (supporting)
Funding acquisition: KN
Investigation: KN (lead) and EI (supporting)
Methodology: KN (lead) and N Kiguchi (supporting)
Project administration: KN
Resources: KN
Supervision: KN (lead), TN (supporting), and N Kiguchi (supporting)
Validation: KN
Visualization: KN
Writing – original draft: KN
Writing – review and editing: KN (lead), TN (supporting), and N Kobayashi (supporting)

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 15 KB - ijmr_v15i1e72281_app1.docx ]

Multimedia Appendix 2
PRISMA-ScR checklist.
[PDF File (Adobe PDF File), 555 KB - ijmr_v15i1e72281_app2.pdf ]

## References

1. Valtorta N, Hanratty B. Loneliness, isolation and the health of older adults: do we need a new research agenda? J R Soc Med 2012;105(12):518-522 [FREE Full text] [doi: 10.1258/jrsm.2012.120128] [Medline: 23288086]
2. Kino S, Stickley A, Arakawa Y, Saito M, Saito T, Kondo N. Social isolation, loneliness, and their correlates in older Japanese adults. Psychogeriatrics 2023;23(3):475-486 [FREE Full text] [doi: 10.1111/psyg.12957] [Medline: 36966745]
3. Dickens AP, Richards SH, Greaves CJ, Campbell JL. Interventions targeting social isolation in older people: a systematic review. BMC Public Health 2011;11:647 [FREE Full text] [doi: 10.1186/1471-2458-11-647] [Medline: 21843337]
4. Steptoe A, Shankar A, Demakakos P, Wardle J. Social isolation, loneliness, and all-cause mortality in older men and women. Proc Natl Acad Sci U S A 2013;110(15):5797-5801 [FREE Full text] [doi: 10.1073/pnas.1219686110] [Medline: 23530191]
5. Tanskanen J, Anttila T. A prospective study of social isolation, loneliness, and mortality in Finland. Am J Public Health 2016;106(11):2042-2048. [doi: 10.2105/AJPH.2016.303431] [Medline: 27631736]
6. Newall NEG, Menec VH. Loneliness and social isolation of older adults: why it is important to examine these social aspects together. J Soc Pers Relat 2017;36(3):925-939. [doi: 10.1177/0265407517749045]
7. Saito M, Shmizu Y, Yamaguchi M, Takei S. Prevalence and characteristics of the socially isolated elderly in a large urban area. Japanese Journal of Social Welfare 2009;50(1):110-122 [FREE Full text]



XSL•FO
RenderX

8. Miura KW, Sekiguchi T, Otake-Matsuura M, Sawada Y, Tanaka E, Watanabe T, et al. Patterns of social relationships among community-dwelling older adults in Japan: latent class analysis. BMC Geriatr 2022;22(75):1-9 [FREE Full text] [doi: 10.1186/s12877-022-02748-7] [Medline: 35078419]
9. Nomura K, Kobayashi N. Gender gaps in strategies for maintaining the social participation and interaction of older adults with people in a local community. Health Soc Care Community 2024;2024(1). [doi: 10.1155/2024/3572577]
10. Wen Z, Peng S, Yang L, Wang H, Liao X, Liang Q, et al. Factors associated with social isolation in older adults: a systematic review and meta-analysis. J Am Med Dir Assoc 2023;24(3):322-330.e6. [doi: 10.1016/j.jamda.2022.11.008] [Medline: 36549651]
11. Tamres LK, Janicki D, Helgeson VS. Sex differences in coping behavior: a meta-analytic review and an examination of relative coping. Pers Soc Psychol Rev 2002;6(1):2-30. [doi: 10.1207/s15327957pspr0601_1]
12. Barreto M, Doyle DM, Maes M. Researching gender and loneliness differently. Ann N Y Acad Sci 2025;1544(1):55-64. [doi: 10.1111/nyas.15283] [Medline: 39761171]
13. Pollak C, Verghese J, Blumen H. Loneliness and functional decline in aging: a systematic review. Res Gerontol Nurs 2023 Jul;16(4):202-212. [doi: 10.3928/19404921-20230503-02] [Medline: 37159388]
14. Haihambo N, Layiwola D, Blank H, Hurlemann R, Scheele D. Loneliness and social conformity: a predictive processing perspective. Ann N Y Acad Sci 2025;1547(1):5-17. [doi: 10.1111/nyas.15324] [Medline: 40173107]
15. Milligan C, Neary D, Payne S, Hanratty B, Irwin P, Dowrick C. Older men and social activity: a scoping review of Men's Sheds and other gendered interventions. Ageing Soc 2015;36(5):895-923 [FREE Full text] [doi: 10.1017/S0144686X14001524]
16. Kelly D, Steiner A, Mason H, Teasdale S. Men's Sheds: a conceptual exploration of the causal pathways for health and well-being. Health Soc Care Community 2019 [FREE Full text] [doi: 10.1111/hsc.12765] [Medline: 31206945]
17. Fakoya OA, McCorry NK, Donnelly M. Loneliness and social isolation interventions for older adults: a scoping review of reviews. BMC Public Health 2020;20(1):129 [FREE Full text] [doi: 10.1186/s12889-020-8251-6] [Medline: 32054474]
18. JBI manual for evidence synthesis. JBI. URL: https://synthesismanual.jbi.global [accessed 2025-06-27]
19. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
20. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
21. Santos-Olmo A, Ausín B, Muñoz M. People over 65 years old in social isolation: description of an effective community intervention in the city of Madrid (Spain). Int J Environ Res Public Health 2022;19(5) [FREE Full text] [doi: 10.3390/ijerph19052665] [Medline: 35270358]
22. Pandya SP. Meditation program mitigates loneliness and promotes wellbeing, life satisfaction and contentment among retired older adults: a two-year follow-up study in four South Asian cities. Aging Ment Health 2021;25(2):286-298. [doi: 10.1080/13607863.2019.1691143] [Medline: 31755300]
23. Mays AM, Kim S, Rosales K, Au T, Rosen S. The leveraging exercise to age in place (LEAP) study: engaging older adults in community-based exercise classes to impact loneliness and social isolation. Am J Geriatr Psychiatry 2021;29(8):777-788 [FREE Full text] [doi: 10.1016/j.jagp.2020.10.006] [Medline: 33268235]
24. Lawlor ER, Cupples ME, Donnelly M, Tully MA. Promoting physical activity among community groups of older women in socio-economically disadvantaged areas: randomised feasibility study. Trials 2019;20(1):234 [FREE Full text] [doi: 10.1186/s13063-019-3312-9] [Medline: 31023329]
25. Alıcı NK, Bahceli PZ, Emiroğlu ON. The preliminary effects of laughter therapy on loneliness and death anxiety among older adults living in nursing homes. Int J Older People Nurs 2018;13(4):e12206. [doi: 10.1111/opn.12206] [Medline: 30004172]
26. Ngiam N, Yee WQ, Teo N, Yow KS, Soundararajan A, Lim JX, et al. Building digital literacy in older adults of low socioeconomic status in Singapore (project wire up): nonrandomized controlled trial. J Med Internet Res 2022;24(12):e40341 [FREE Full text] [doi: 10.2196/40341] [Medline: 36459398]
27. Dodge H, Yu K, Wu C, Pruitt P, Asgari M, Kaye JA, et al. Internet-based conversational engagement randomized controlled clinical trial (I-CONECT) among socially isolated adults 75+ years old with normal cognition or mild cognitive impairment: topline results. Gerontologist 2024 Apr 01;64(4) [FREE Full text] [doi: 10.1093/geront/gnad147] [Medline: 37935416]
28. Dodge HH, Zhu J, Mattek NC, Bowman M, Ybarra O, Wild KV, et al. Web-enabled conversational interactions as a means to improve cognitive functions: results of a 6-week randomized controlled trial. Alzheimers Dement (N Y) 2015 May;1(1):1-12 [FREE Full text] [doi: 10.1016/j.trci.2015.01.001] [Medline: 26203461]
29. Zamir S, Hennessy C, Taylor A, Jones R. Intergroup 'Skype' quiz sessions in care homes to reduce loneliness and social isolation in older people. Geriatrics (Basel) 2020 Nov 11;5(4) [FREE Full text] [doi: 10.3390/geriatrics5040090] [Medline: 33187242]
30. Hind D, Mountain G, Gossage-Worrall R, Walters SJ, Duncan R, Newbould L, et al. Putting Life in Years (PLINY): a randomised controlled trial and mixed-methods process evaluation of a telephone friendship intervention to improve mental well-being in independently living older people. Public Health Research 2014:1-225. [Medline: 25642518]




XSL•FO
RenderX

31. Czaja S, Boot W, Charness N, Rogers W, Sharit J. Improving social support for older adults through technology: findings from the PRISM randomized controlled trial. Gerontologist 2018;58(3):467-477 [FREE Full text] [doi: 10.1093/geront/gnw249] [Medline: 28201730]
32. Sen K, Prybutok G, Prybutok V. The use of digital technology for social wellbeing reduces social isolation in older adults: a systematic review. SSM Popul Health 2022 Mar;17 [FREE Full text] [doi: 10.1016/j.ssmph.2021.101020] [Medline: 35024424]
33. Chen YR, Schulz PJ. The effect of information communication technology interventions on reducing social isolation in the elderly: a systematic review. J Med Internet Res 2016;18(1):e18 [FREE Full text] [doi: 10.2196/jmir.4596] [Medline: 26822073]
34. Jarvis M, Padmanabhanunni A, Balakrishna Y, Chipps J. The effectiveness of interventions addressing loneliness in older persons: an umbrella review. Int J Afr Nurs Sci 2020 May 17;12:100177-100118 [FREE Full text] [doi: 10.1016/j.ijans.2019.100177]
35. van de Venter E, Buller A. Arts on referral interventions: a mixed-methods study investigating factors associated with differential changes in mental well-being. J Public Health (Oxf) 2015;37(1):143-150. [doi: 10.1093/pubmed/fdu028] [Medline: 24839293]
36. Shepherd G, Astbury E, Cooper E, Dobrzynska W, Goddard E, Murphy H. The challenges preventing men from seeking counselling or psychotherapy. Ment Health Prev 2023:1-10 [FREE Full text] [doi: 10.1016/j.mhp.2023.200287]
37. Meehan DE, Grunseit A, Condie J, HaGani N, Merom D. Social-ecological factors influencing loneliness and social isolation in older people: a scoping review. BMC Geriatr 2023 Nov 09;23(1):726 [FREE Full text] [doi: 10.1186/s12877-023-04418-8] [Medline: 37946155]
38. You are not alone. The Office for Policy on Loneliness and Isolation. URL: https://www.notalone-cao.go.jp/english/ [accessed 2025-08-21]


## Abbreviations

**JBI:** Joanna Briggs Institute
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews


*Edited by N Cahill, M Balcarras; submitted 07.Feb.2025; peer-reviewed by C Chikwe, VS Silva, J Scott; comments to author 14.Aug.2025; accepted 29.Dec.2025; published 18.Feb.2026.*

*Please cite as:*
*Nomura K, Kiguchi N, Inomata E, Nakamachi T, Kobayashi N*
*Treatment of Gender in Research on Intervention Programs Targeting Social Isolation and Loneliness Among Older Adults: Scoping Review*
*Interact J Med Res 2026;15:e72281*
*URL: https://www.i-jmr.org/2026/1/e72281*
*doi:10.2196/72281*
*PMID:*


©Kenta Nomura, Naoto Kiguchi, Eisuke Inomata, Takeshi Nakamachi, Norikazu Kobayashi. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 18.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Review

# Internet Use and Self-Rated Health Among Older Adults: Scoping Review

Hanan AboJabel[1], PhD; Fareeda Abo-Rass[2], PhD

[1]The Paul Baerwald School of Social Work and Social Welfare, Hebrew University of Jerusalem, Jerusalem, Israel

[2]Department of Social Work, Ben-Gurion University of the Negev, Beer-Sheva, Israel

**Corresponding Author:**
Hanan AboJabel, PhD
The Paul Baerwald School of Social Work and Social Welfare
Hebrew University of Jerusalem
Mount Scopus Campus
Jerusalem, 9190501
Israel
Phone: 972 0556810001
Email: hanan.abo-jabel@mail.huji.ac.il

## Abstract

**Background:** Self-rated health (SRH) is a robust predictor of morbidity, functional decline, and mortality in later life. As internet use becomes increasingly embedded in older adults' daily routines, clarifying its association with SRH and the pathways through which it may operate is important for research, practice, and policy.

**Objective:** This scoping review aimed to map and characterize the international evidence on the association between internet use and SRH among older adults, synthesize how potential mediators and moderators have been examined, and identify key methodological, theoretical, and population gaps in the literature.

**Methods:** Guided by the Joanna Briggs Institute methodology and PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) reporting standards, we conducted a scoping review and searched 5 databases: PubMed, CINAHL, AgeLine, PsycINFO, and Web of Science. The final search was performed on February 5, 2024. Reference lists were screened, and Google Scholar searches were conducted as supplementary search methods.

**Results:** Database searches identified 4294 records; after removing 615 duplicates, 3679 records were screened, and 77 full texts were assessed, resulting in 27 included studies. All included studies were quantitative, and the evidence base was predominantly cross-sectional (25/27). Explicit theoretical frameworks were used in 6 out of 27 studies. Most studies were published between 2019 and 2024 (22/27) and were conducted most frequently in China (11/27) and the United States (7/27). All studies were conducted in high-income countries. SRH was typically assessed using a single-item measure, while internet use was operationalized as access/use (yes/no), frequency, and/or purpose/domain-specific measures. Most studies reported a statistically significant positive association between internet use and better SRH (24/27), with socially oriented uses (eg, communication and social participation) showing the most consistent associations. Mediating pathways were examined in 6 out of 27 studies, and most often suggested social mechanisms such as greater social support, higher social engagement, and lower loneliness. Subgroup heterogeneity was reported in 10 out of 27 studies, including differences by age, gender, residence, and marital status.

**Conclusions:** Overall, internet use, particularly socially oriented use, was most consistently associated with better SRH among older adults. Policy efforts should support digital inclusion by improving access, skills, and ongoing assistance that enable meaningful use for social connection and service access. At the same time, nondigital options are essential to avoid excluding older adults who do not use the internet. In addition, evidence gaps, including limited use of theoretical frameworks and scarce data from low- and middle-income countries, underscore the need for theory-informed longitudinal and intervention studies to strengthen causal inference, expand knowledge on mediating and moderating factors, and assess generalizability across diverse contexts.

***(Interact J Med Res 2026;15:e76930)*** doi:10.2196/76930

**KEYWORDS**

internet use; older adults; scoping review; self-rated health; social participation



XSL•FO
RenderX

## Introduction

Self-rated health (SRH) is a measure of individuals' subjective assessments of their general health, encompassing physical, mental, social, and cultural aspects [1-3]. Usually, SRH is assessed using a single item, often phrased as: "Would you say your general health is excellent, very good, good, fair, or poor?" [4]. This scale is widely used to assess health status among older adults, as it is considered valid, reliable, quick, and easy to administer [5,6]. Indeed, SRH is regarded as an important predictor of healthy aging, future morbidity, functional capacity, hospitalization, health care consumption, and mortality [7-11]. The literature indicates that a variety of factors may play an important role in good SRH among older adults, including background characteristics (eg, younger age, higher education, and male gender), social support networks, good mental and physical health, the ability to care for oneself, and healthy behaviors (eg, sufficient physical activity) [1,12-17].

An additional factor that may be related to older adults' SRH is internet use. The internet offers **a range of opportunities** to promote independent living among older adults and to improve their quality of life and overall well-being [18-21]. It facilitates communication and helps older adults maintain relationships with family and friends, as well as establish new social connections [19,21]. In addition, the internet enables participation in online leisure activities [22,23]. Furthermore, older adults can use the internet to seek information, including health-related information, potentially increasing health literacy, and supporting better health-related decision-making [24-26]. However, extensive internet use also entails potential risks. For instance, frequent internet use for nonsocial purposes, such as information seeking and entertainment, may reduce social interactions and weaken existing social networks, thereby increasing social exclusion [27-29]. There is also a risk of internet addiction among older adults [30], along with exposure to privacy invasion and data theft [31,32]. These opportunities and risks may, in turn, influence older adults' SRH.

In recent years, a growing body of research has examined the association between internet use and SRH among older adults; however, this literature has not yet been comprehensively synthesized and organized. Conducting a scoping review is therefore warranted, given the increasing integration of the internet into later life and the expansion of online services, especially in health and social domains [33-35]. In addition, digital technologies have been suggested as potentially contributing to responses to population aging and to alleviating pressures on health systems [36,37].

Therefore, the aim of this scoping review is to map and synthesize the existing international literature on internet use and SRH in later life by (1) describing how internet use and SRH have been conceptualized, operationalized, and examined across studies, as well as mapping the reported patterns of findings regarding the direction and nature of the association between internet use and SRH; (2) mapping how the literature addresses potential pathways linking internet use and SRH, including mediating and moderating factors, where mediators are defined as variables that explain the mechanisms through which one variable influences another, and moderators are defined as variables that affect the strength or direction of this relationship [38]; and (3) identifying methodological and theoretical gaps, as well as underrepresented populations within the existing evidence base. Identifying these elements is essential for informing the development of evidence-based interventions aimed at promoting SRH and well-being in later life through the use of digital technologies, as well as for supporting the design and advancement of policies in the fields of aging and digital health.

## Methods

### Overview

This scoping review was conducted in accordance with the Joanna Briggs Institute (JBI) methodology for scoping reviews and is reported following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [39]. In line with scoping review methodology, no formal appraisal of methodological quality was conducted [39]. An overview of adherence to the PRISMA-ScR reporting criteria is presented in Multimedia Appendix 1.

### Search Strategy

The search strategy was reported in accordance with the PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Searching) guideline to enhance transparency and reproducibility [40]. An overview of adherence to the PRISMA-S reporting criteria is provided in Multimedia Appendix 2. The search strategy was developed by the first author, who has expertise in aging and technology, and reviewed by the second author. It was developed iteratively, guided by the review objectives, key concepts identified in the literature, and preliminary exploratory searches. The search was conducted across 5 databases relevant to the research topic: PubMed, CINAHL, AgeLine, PsycINFO, and Web of Science. Searches were limited to peer-reviewed journal articles published in English. The final search was performed on February 5, 2024, and included all publications available up to that date. The search strategy combined keywords and Boolean operators representing 3 core concepts: internet use ("Internet" or "web" or "online" or "digital" or "technology" or "information and communication technology" or "social media" or "computer" or "smartphone" or "tablet"), SRH ("self-rated health" or "self-reported health" or "self-assessed health" or "perceived health" or "subjective health"), and older adults ("older adults" or "aging" or "aged" or "older people" or "elderly" or "seniors"). Full search strategies for each database are presented in Multimedia Appendix 3. To enhance the comprehensiveness of the search, reference lists of all included articles were screened, and Google Scholar was used as a supplementary search tool, consistent with recommendations to use Google Scholar as a complementary resource in systematic and scoping reviews rather than as a stand-alone database [41]. No additional eligible studies were identified through these supplementary searches. Duplicates were removed manually by the first author by comparing titles, authors, and



XSL•FO
RenderX

publication details across records, and the deduplication was verified by the second author.

### Inclusion and Exclusion Criteria

Studies were eligible for inclusion if they met the following criteria: (1) the study reported original empirical findings derived from quantitative, qualitative, or mixed methods research, based on either primary data collection or secondary data analysis; (2) the study explicitly addressed the conceptualization and/or operationalization of the key constructs and examined the relationship between internet use and SRH; (3) the study population consisted primarily of older adults, defined as individuals aged 60 years and older, or was explicitly described by the original authors as comprising older adults, in line with commonly used definitions in previous systematic and scoping reviews (Huang et al [42]). Studies with broader age ranges were also eligible if results for older adults were reported separately; (4) the article was published in English; and (5) the article appeared in a peer-reviewed journal. Studies were excluded if they were opinion pieces, conference abstracts, book reviews, book chapters, theses, or doctoral dissertations. In addition, studies in which internet use was modeled as the outcome variable and SRH as the independent variable were excluded. This decision ensured methodological consistency by maintaining a uniform analytical direction, as is common in systematic and scoping reviews [43,44], thereby facilitating more coherent evidence mapping and synthesis across studies.

### Data Charting

Data charting was guided by a structured framework informed by the JBI data extraction template for scoping reviews and adapted to the objectives of this review [45]. Data were charted from each included article at the individual-study level, including bibliographic details, study aims, study design, stated theoretical framework, sample characteristics, and the measures used to assess internet use and SRH. Key findings were also charted, including the reported association between internet use and SRH and mediators or moderators explicitly examined or discussed by the study authors. The first author conducted the data charting manually by reviewing each included article in full. To enhance reliability, the second author reviewed the charted data and the data charting framework for accuracy and consistency, and uncertainties were resolved through discussion and consensus.

### Data Analysis and Synthesis

Prior to finalizing the approach for analysis and presentation, the authors conducted a preliminary review of 5 randomly selected included articles to inform the structure and focus of the subsequent analysis. Building on the charted data, study characteristics were organized descriptively to map the evidence base and to support a systematic presentation of study characteristics in the Results section (Table 1). Subsequently, a narrative synthesis of the reported findings was conducted to identify recurring patterns across studies. The synthesis summarized the direction of the association between internet use and SRH and organized results according to key dimensions aligned with the review objectives, including the type of internet use and variation across population subgroups. Findings related to mediating pathways and between-study variation were integrated where reported, to support a coherent presentation of results in the Results section, while Table 2 provided a concise overview of the synthesized evidence. To ensure the reliability of the analytical process, both authors jointly reviewed the organization and presentation of the synthesized findings, and any discrepancies were resolved through discussion and consensus.



XSL•FO
RenderX

**Table 1.** Characteristics of included studies.

| Authors (country) | Aims | Study design/theoretical framework | Sample characteristics | Measure of SRH[a] | Measure of internet use |
|---|---|---|---|---|---|
| Chen et al 2022 [46] (China) | To examine the association between internet use and health outcomes, including the mediating role of cultural engagement | Cross-sectional; Chinese General Social Survey (2015 and 2017 waves)<br>Activity theory (Fernández-Ballesteros et al 2021 [47]) | N=6066 older adults aged ≥60 years; mean age 69.19 (SD 7.27) years; 52% women; mean education 1.5492 (SD 0.6378); range 1 (“primary school”) to 3 (“college or above”) | Single item (unhealthy/healthy) | Internet use (yes/no) |
| Chopik 2016 [48] (United States) | To examine the association between socially oriented technology use and health-related outcomes | Cross-sectional; Health and Retirement Study (2012 wave) | N=591 older adults; mean age 68.18 (SD 10.75) years; 55.5% women; mean education 13.28 (SD 2.70) years; 73.9% White, 19.9% Black/African American, 6.1% Hispanic | Single item (5-point scale: 1 “poor” to 5 “excellent”) | Social technology use was measured using 5 modalities (email, social networking sites, online video/phone calls, online chat/instant messaging, smartphone); summed index (0-5) |
| Ding et al 2023 [49] (China) | To examine the association between internet access and health-related outcomes | Cross-sectional; China Health and Retirement Longitudinal Study (2011, 2013, 2015, and 2018 waves) | N=57,960 older adults; mean age 58.7 (SD 8.09) years; 50% women; mean education 1.712 (SD 1.304) | Single item (poor health/good health) | Internet access (yes/no) |
| Duplaga et al 2021 [50] (Poland) | To examine the association between internet use and health-related outcomes | Cross-sectional; telephone survey | N=1000 adults aged ≥50 years; 39.3% aged 50-59 years, 28.5% aged 60-69 years, 29.7% aged ≥70 years; 55.8% women; education: 27.8% below secondary, 45.9% secondary/postsecondary, 26.3% university | Single item (“unsatisfactory”/”at least satisfactory”) | Frequency of internet use (4-point scale: 0 “no use,” 1 “a few times a month or less,” 2 “a few times a week,” 3 “every day”) |
| Falk Erhag et al 2019 [51] (Sweden) | To examine the association between internet use and SRH | Cross-sectional; Gothenburg H70 Birth Cohort Study (2014-2016 wave); personal interview and self-administered questionnaire | N=1136 older adults aged 70 years; 69.3% men; 82.9% had >9 years of education; 28.5% had a university degree | Single item (5-point scale: “excellent” to “poor”) | Frequency of internet use (7-point scale: “never” to “daily or several times a day”) |
| Fjell et al 2020 [52] (Norway) | To examine the association between SRH and factors related to demographics, lifestyle, health conditions, and medical diagnoses | Cross-sectional; face-to-face interview | N=233 older adults participating in a preventive home visit program; mean age 77 (range 75-79) years; 50% women; education: 25% >7 years, 75% ≤7 years | Single item (5-point scale: 1 “poor” to 5 “excellent”) | Internet use (yes/no) |
| Gracia and Herrero 2009 [53] (Spain) | To examine the association between internet use and SRH | Cross-sectional; telephone/online survey (Digital divide and quality of life among older people) | N=709 adults aged ≥55 years; 56.7% aged 55-64 years, 43.3% aged 65-74 years; 52.2% women; social class: 8.1% high, 9.2% medium-high, 26.3% medium, 35.3% medium-low, 21% low | Single item (good health/poor health) | Internet use (user/nonuser) |
| Jeon and Choi 2024 [54] (Korea) | To examine the association between internet use and physical and psychological health outcomes | Cross-sectional; Living Profiles of Older People Survey (2020 wave); face-to-face interviews | N=5094 older adults aged ≥65 years; mean age 70.71 (SD 2.8) years; 54.1% women; education: 9.7% college+, 42.6% high school, 25.3% middle school, 19.3% elementary school, 3.1% no formal education | Single item (5-point scale: 1 “very good” to 5 “very poor”) | Internet use:<br>Interpersonal communication: 3 yes/no items; summed index (0-3)<br>Instrumental use: 8 yes/no items; summed index (0-8) |
| Kim et al 2020 [55] (United States) | To examine the association between ICT[b] use and health-related outcomes | Cross-sectional; National Health and Aging Trends Study (2010 wave) | N=4976 older adults aged ≥65 years; mean age 73.7 (SD 0.09) years; 54.3% women; mean education 2.8 (SD 0.4); range 1 (“less than 12th grade”) to 5 (“graduate degree”); 83.5% White, 7.6% Black, 2.9% other | Single item (5-point scale: 1 “poor health” to 5 “excellent health”) | ICT:<br>IT (6 items; yes/no): 3 items for personal tasks >3 items for health-related information.<br>Communication technology (1 item; frequency): sending messages by email or texting (rarely/some days/most days). |



XSL•FO
RenderX

| Authors (country) | Aims | Study design/theoretical framework | Sample characteristics | Measure of SRH[a] | Measure of internet use |
|---|---|---|---|---|---|
| Koopman-Boyden and Reid 2009 [56] (New Zealand) | To examine the association between internet use and well-being | Cross-sectional; phone/online survey | N=1680 older adults aged 65-84 years; 31% aged 65-69 years, 28% aged 70-74 years, 25% aged 75-79 years, 16% aged 80-84 years | Single item (satisfied/dissatisfied with health) | Internet use (yes/no) |
| Lee and Jang 2022 [57] (Korea) | To examine the association between changes in internet use during the COVID-19 pandemic and SRH, comparing young-old (65-74) and old-old (≥75) adults | Cross-sectional; Digital Divide Survey (2020 wave); face-to-face interviews | N=1150 older adults aged ≥65 years: young-old (n=670; mean age 69.18 (SD 2.8) years; 52.5% women; 23.4% <middle school, 76.6% ≥high school) and old-old (n=480; mean age 79.93 (SD) 3.6 years; 62.9% women; 44.8% <high school, 55.2% ≥high school) | Single item (4-point scale: 1 “not satisfied at all” to 4 “very satisfied”) | Internet use: 4 domains (each rated on a 5-point scale from 1 “significantly decreased” to 5 “significantly increased”): (1) social networking and information-sharing services; (2) social participation services; (3) daily services; and (4) search, email, and content services |
| Lee et al 2018 [58] (United States) | To examine the association between ICT use and health-related outcomes among older cancer survivors | Cross-sectional; National Health and Aging Trends Study (2011 wave) | N=1411 community-dwelling cancer survivors aged ≥65; 53.6% aged 65-74 years, 36.2% aged 75-84 years, 10.2% aged ≥85 years; 48% women; mean education 2.9 (SD 0.22); range 1 (“less than 12th grade”) to 5 (“graduate degree”); 90% White, non-Hispanic | Single item (5-point scale: 1 “poor” to 5 “excellent”) | ICT use:<br>Communication technology: Frequency of emailing/texting (no or rarely/some days/most days)<br>Information technology: 6 yes/no items, summarized as 2 indices: personal tasks (3 items) and health-related internet use (3 items) |
| Li et al 2023 [59] (China) | To examine the association between online and offline social activities and health-related outcomes | Cross-sectional; World Values Survey (wave 7, China sample) and China Health and Retirement Longitudinal Study (2018 wave) | Sample 1: N=598 older adults aged ≥60 years; mean age 64.76 (SD 2.97) years; 52.7% women; education: 53.8% primary school or below, 26.4% middle school, 15.6% high school/vocational, 3.8% bachelor’s, 0.3% master’s or above. Sample 2: N=9434 older adults; mean age 68.53 (SD 6.44) years; 50% women; education: 73.7% primary school or below, 16.3% middle school, 8.5% high school/vocational, 0.7% associate degree, 0.6% bachelor’s, 0.03% master’s or above | Single item (5-point scale: 1 “very poor” to 5 “very good”) in both samples | Sample 1: Frequency of internet use (5-point scale: 1 “never” to 5 “daily”). Sample 2: Frequency of internet use (0-3 scale: 0 “no participation” to 3 “almost daily”; past month) |
| Liu et al 2023 [60] (China) | To examine the association between internet use and SRH, including the mediating role of social engagement and heterogeneity by living arrangements | Cross-sectional; China Health and Retirement Longitudinal Study (2018 wave) | N=14,587 middle-aged and older adults aged ≥45 years; mean age 61 (SD 9.248) years; 52% men; education: 62% elementary school or below, 23% junior high, 13% high school or above | Single item (5-point scale: 1 “very poor” to 5 “very good”) | Two measures:<br>Internet use (yes/no)<br>Frequency of internet use (4-point scale: 0 “never” to 3 “almost every day”) |
| Liu et al 2022 [61] (China) | To examine the association between internet use and SRH, including the mediating role of social support | Cross-sectional; Chinese General Social Survey (2017 wave) | N=4234 older adults aged ≥60 years; mean age 69.34 (SD 7.37) years; 48% men; mean education 6.65 (SD 4.61) years | Single item (5-point scale: 1 “very unhealthy” to 5 “very healthy”) | Internet access (yes/no) |
| Lyu and Sun 2021 [62] (China) | To examine the association between internet use and SRH, including the mediating role of social capital | Cross-sectional; China Family Panel Studies (2018 wave) | N=7193 older adults aged ≥60 years; age distribution: 64.83% aged 60-69 years, 28.69% aged 70-79 years, 6.48% aged ≥80 years; 51.01% men; education: 88.39% had 0-9 years, 10.86% had 10-15 years, 0.75% had ≥16 years | Single item (unhealthy/healthy) | Internet use (yes/no) |



XSL•FO
RenderX

| Authors (country) | Aims | Study design/theoretical framework | Sample characteristics | Measure of SRH[a] | Measure of internet use |
|---|---|---|---|---|---|
| Millar et al 2020 [63] (United States) | To examine the association between problem-solving in technology-rich environments and SRH, focusing on age differences | Cross-sectional; Program for the International Assessment of Adult Competencies (2012-2014 wave)<br>Paasche-Orlow and Wolf's [64] model (2007) and Gewald and Rockmann's [65] model (2016) | N=3260 adults aged ≥35 years; age: 26.65% 35-44 years, 28.71% 45-54 years, 27.12% 55-65 years, 16.58% ≥66 years; 54.82% women; education: 60.74% high school or less, 39.25% college or higher | Single item (good health/poor health) | Internet use for health information seeking (yes/no) |
| Nakagomi et al 2022 [66] (Japan) | To examine the association between internet use and subsequent health and well-being | Longitudinal; Japan Gerontological Evaluation Study (3 waves: 2013, 2016, 2019); self-administered questionnaires | N=3903 older adults aged ≥65; mean age 71.34 (SD 4.65) years; 62.97% women; education: 68.07% had ≥10 years, 31.92% had ≤9 years | Single item (excellent/good vs others) | Frequency of internet use (4-point scale: not at all/a few times a month/a few times a week/almost every day) |
| Sims et al 2016 [67] (United States) | To examine the association between ICT use and well-being among the oldest-old | Cross-sectional; online/telephone survey<br>Socioemotional selectivity theory and biological models of aging | N=445 older adults aged ≥80 years; mean age 84 (SD 3) years; 64% women; 45% had more than high school education; 26% non-White | Single item (5-point scale: 1 "excellent" to 5 "poor") | ICT use: number of ICT devices/applications used (3-point scale: 0 "none," 1 "one," 2 "two or more"); motivation for ICT use: 2 items (connect with family/friends; learn new information/skills), each rated on a 5-point scale (1 "strongly disagree" to 5 "strongly agree") |
| Swed et al 2020 [68] (United States) | To examine the association between internet use and SRH among older military veterans | Cross-sectional; National Survey of Veterans (2010 wave); mail questionnaire | N=8539 older adults aged ≥60 years; mean age 61.02 (SD 0.02) years; 91.5% men; education: 5.4% <high school, 26% high school, 30% some college, 27% college degree, 11.7% professional degree; 80.6% White, 10.9% Black, 4.8% Hispanic, 3.7% other | Single item (3-point scale: "fair/poor," "good/very good," "excellent") | Frequency of internet use (4-point scale: daily/once a week, not daily/once a month-once per year/does not use the internet) |
| Tavares 2020 [69] (17 European countries and Israel) | To examine the association between internet use and SRH, and whether this association varies across countries with different levels of eHealth policy development | Cross-sectional; Survey of Health, Ageing and Retirement in Europe (SHARE), wave 6<br>Social ecological model (Dahlgren and Whitehead 1991 [70]) | N=66,279 adults aged ≥50 years from 18 countries; mean age 67.93 years; 46.8% men; mean education 10.83 years | Single item (5-point scale: 1 "poor health" to 5 "excellent health") | Internet use in the past 7 days (yes/no) |
| Wang et al 2020 [71] (China) | To examine the association between internet use and health-related outcomes and whether individual cognitive ability moderates this association | Longitudinal; Chinese General Social Survey (2012 and 2015 waves) | 2012 wave: N=2821 older adults aged ≥60 years; mean age 69.10 (SD 7.20) years; 56% men; mean education 2.40 (SD 1.27); range 1 ("no education") to 7 ("graduate and above"). 2015 wave: N=3185 older adults aged ≥60 years; mean age 69.39 (SD 7.46) years; 49% men; mean education 2.39 (SD 1.26); range 1 ("no education") to 7 ("graduate and above") | Single item (5-point scale: 1 "relatively poor" to 5 "very healthy") | Frequency of internet use (5-point scale: 1 "never" to 5 "very frequently") |



XSL•FO
RenderX

| Authors (country) | Aims | Study design/theoretical framework | Sample characteristics | Measure of SRH[a] | Measure of internet use |
|---|---|---|---|---|---|
| Wei and Guo 2023 [72] (China) | To examine the association between smartphone use and health-related outcomes | Cross-sectional; survey; face-to-face interviews | N=1110 older adults aged ≥55 years; age category (coded): mean 2.27 (SD 0.73); range 1 ("55-64"), 2 ("65-74"), 3 (">75"); 47.02% men; mean education 1.71 (SD 0.91); range 1 ("primary school or below") to 4 ("junior high or above"); 44% lived in urban areas | Single item (3-point scale: 1 "poor," 2 "average," 3 "good") | Smartphone use (yes/no) |
| Wei et al 2022 [73] (China) | To examine the association between WeChat use and SRH | Cross-sectional; China Health and Retirement Longitudinal Study (2018 wave)<br>Social participation theory (Lian et al 1999 [74]) | N=5442 older adults aged ≥60 years; mean age 69.34 (SD 7.23) years; 51% women; mean education 3.00 (SD 1.91); range 1 ("illiterate") to 11 ("PhD") | Single item (5-point scale: 1 "very good" to 5 "very bad") | WeChat use (use/do not use) |
| Wen et al 2023 [75] (China) | To examine the association between internet use and health-related outcomes | Cross-sectional; China Health and Retirement Longitudinal Study (2018 wave) | N=13,474 middle-aged and older adults; mean age 61.50 (SD 9.30) years; 51.57% women; education: 87.45% below middle school, 10.47% high school/vocational training, 1.81% above high school; 37.39% lived in urban areas | Single item (3-point scale: 1 "positive," 2 "general," 3 "negative") | Internet use (yes/no) |
| Xiaobing and Meng 2022 [76] (China) | To examine the association between internet use and community participation, including the mediating role of SRH | Cross-sectional; China Health and Retirement Longitudinal Study (2016 wave)<br>Displacement theory | N=8856 older adults aged ≥60 years; mean age 70.20 (SD 7.55) years; 51.63% women; mean education 3.29 (SD 1.31); range 1 ("illiterate") to 6 ("above college") | Single item (5-point scale: "very unhealthy" to "healthy") | Frequent internet access (yes/no)<br>Frequency of internet use (5-point scale: 0 "rarely" to 4 "always")<br>Internet as a primary source of information (yes/no) |

[a]SRH: self-rated health.

[b]ICT: information and communication technology.



XSL•FO
RenderX

**Table 2.** Main results of included studies on internet use and self-rated health (SRH).

| Authors (country) | Level of SRH | Level of internet usage | The impact of internet use on SRH | |
|---|---|---|---|---|
| | | | Positive/negative/NS[a] | Mediators, moderators, and heterogeneity across groups |
| Chen et al 2022 [46] (China) | 39% reported being healthy | 22% used the internet | Positive | Mediated by cultural engagement |
| Chopik 2016 [48] (USA) | Moderate-high level; mean 3.26, SD 1.05 | Low level (social technology use for social connection); mean 1.37, SD 1.42 | Positive | Mediated by reduced loneliness |
| Ding et al 2023 [49] (China) | 74% reported good health | 24% had internet access | Positive | Stronger among vulnerable groups (eg, difficulties in ADL[b], no social participation, or no spouse) |
| Duplaga 2021 [50] (Poland) | 67.9% reported at least satisfactory health | 51.1% did not use the internet, 7.9% used the internet a few times a month or less, 14.8% used the internet a few times a week, and 26.2% used the internet every day | Negative | N/A[c] |
| Falk Erhag et al 2019 [51] (Sweden) | 11.97% excellent; 36.18% very good; 35.56% good; 14.35% moderate; 1.94% poor | 66.10% reported using the internet daily | Positive (minor effect compared to health-related variables) | N/A |
| Fjell et al 2020 [52] (Norway) | Moderate level; mean 2.84, SD 0.88 | 44% used the internet | Positive | N/A |
| Gracia and Herrero 2009 [53] (Spain) | 83.2% reported good health | 17.3% used the internet | Positive (the relationship between internet use and SRH disappeared once social class was considered) | N/A |
| Jeon and Choi 2024 [54] (Korea) | Moderate level; mean 2.44, SD 0.81 | Internet use for interpersonal communication: high level; mean 2.24, SD 0.74<br>Internet use for instrumental purposes: moderate level; mean 3.22, SD 2.22 | Positive (the impact of internet use for instrumental purposes was stronger than the impact of internet use for communication purposes on SRH) | N/A |
| Kim et al 2020 [55] (USA) | Moderate-high; mean 3.4, SD 0.02 | Information technology—personal tasks: 32% used the internet for personal tasks; 44% shopped groceries/personal items; 44.2% paid bills/banking; 17.6% ordered/refilled prescriptions<br>Information technology—health-related information: 15.9% contacted medical providers; 11.8% handled medical/health insurance matters; 36.2% gathered information about health conditions<br>Communication technology: 65.5% emailed/texted most days; 28.4% some days; 15.1% rarely | Positive for overall ICT[d]; NS for information technology only; NS for communication technology only. | N/A |
| Koopman-Boyden and Reid 2009 [56] (New Zealand) | N/A | 51% used the internet | Positive | N/A |



XSL•FO
RenderX

| Authors (country) | Level of SRH | Level of internet usage | The impact of internet use on SRH | |
|---|---|---|---|---|
| | | | Positive/negative/NS[a] | Mediators, moderators, and heterogeneity across groups |
| Lee and Jang 2022 [57] (Korea) | Young-old: moderate-high level; mean 2.61, SD 0.80<br>Old-old: moderate-high level; mean 2.44, SD 0.70 | Young-old: moderate to moderate-high level of increase in internet use: social networking and information-sharing services (mean 3.24, SD 0.58), social participation (mean 2.89, SD 0.59), daily services (mean 3.20, SD 0.59), search/email/content services (mean 3.26, SD 0.57)<br>Old-old: moderate level of increase in internet use: social networking and information-sharing services (mean 2.98, SD 0.44), social participation (mean 2.87, SD 0.51), daily services (mean 2.98, SD 0.42), search/email/content services (mean 3.01, SD 0.42) | Positive | Age differences: positive association between SRH and social networking/information-sharing services and online daily services in both groups; social participation services and search/email/content services were positively associated with SRH only among the old-old group |
| Lee et al 2018 [58] (United States) | Moderate-high level; mean 3.2, SD 0.03 | Communication technology: 15.7% emailed/texted some days; 26.8% most days<br>Information technology (personal tasks): 31.7% used the internet for personal tasks<br>Health matters: 26.5% used the internet for health matters | Positive for communication technology (significant only when used most days); NS for information technology and health matters | N/A |
| Li et al 2023 [59] (China) | N/A | Sample 1: moderate level of frequency of internet use; mean 3.03, SD 1.87<br>Sample 2: mean 0.19, SD 0.71; 6.7% used the internet in the last month | Positive among both samples | N/A |
| Liu et al 2023 [60] (China) | Moderate-high level; mean 3.079, SD 1.021 | 14% used the internet<br>Low frequency of internet usage; mean 0.387, SD 0.982 | Positive | Mediated by social engagement; stronger contribution among those with children (vs without children) |
| Liu et al 2022 [61] (China) | Moderate level; mean 2.97, SD 1.07 | 23.1% had access to the internet | Positive | Mediated by social support (relatives and friends); stronger among male older adults, younger older adults, and rural older adults |
| Lyu and Sun 2021 [62] (China) | 53.47% reported that they were healthy | 12.87% used the internet | Positive | Mediated by social capital |
| Millar et al 2020 [63] (United States) | 85.2% reported good health | 69.96% used the internet to seek health information | NS | N/A |
| Nakagomi et al 2022 [66] (Japan) | 87.8% reported excellent/good health | 47.7% not at all; 14.7% a few times a month; 13.5% a few times a week; 24.2% almost every day | Positive (modest link; significant only for almost every day internet use); NS for a few times a month/a few times a week | N/A |
| Sims et al 2016 [67] (United States) | Moderate level; mean 2.87, SD 0.87 | Mean number of devices/applications used 1.23, SD 0.83<br>Moderate-high level of using technology to connect with family and friends; mean 3.66, SD 1.28<br>Moderate level of using technology to learn new information and skills; mean 2.61, SD 1.44 | Positive | Mediated by using technology to learn new information; not mediated by using technology to connect with family and friends |
| Swed et al 2020 [68] (United States) | 72.4% had good/very good/excellent health | 48.7% used the internet daily; 16.6% used it once a week (not daily); 5.6% used it once a month-once per year; 29.1% did not use the internet | Positive | N/A |



XSL•FO
RenderX

| Authors (country) | Level of SRH | Level of internet usage | The impact of internet use on SRH | |
|---|---|---|---|---|
| | | | Positive/negative/NS[a] | Mediators, moderators, and heterogeneity across groups |
| Tavares 2020 [69] (17 European countries and Israel) | 11.42% poor; 28.56% fair; 35.61% good; 17.62% very good; 6.78% excellent | 48.4% used the internet in the previous 7 days | Positive | Stronger in less eHealth-developed countries; not beneficial for older adults with low levels of health |
| Wang et al 2020 [71] (China) | Moderate-high level in both waves; mean 3.01, 3.18, SD 1.05, 1.06 (2012 and 2015, respectively) | Low level in both waves; mean 1.20, SD 0.78 and mean 1.35, SD 0.93 (2012 and 2015, respectively) | NS | Not moderated by individual cognitive ability |
| Wei and Guo 2023 [72] (China) | Moderate-high level; mean 2.377, SD 0.760 | 57% used a smartphone | Positive | Stronger among urban residents and among adults aged >75 years; mediated by performance expectations and individualized needs |
| Wei et al 2022 [73] (China) | Moderate-high level; mean 3.12, SD 1.04 | 5% used WeChat | Positive | Stronger among younger-older adults (<70 years); positive effects among both older men and women |
| Wen et al 2023 [75] (China) | 23.85% positive health; 49.49% general health; 26.65% negative health | 24.32% used the internet | Positive | Significant among older adults living in both rural and urban settings |
| Xiaobing and Meng 2022 [76] (China) | 8.34% very unhealthy; 39.08% relatively unhealthy; 34.16% average; 14.72% relatively healthy; 3.69% very healthy | Internet use: 88.91% did not use the internet; internet use frequency: 83.68% never; 4.98% rarely; 4.93% sometimes; 4.43% often; 1.98% always**;** internet as an information source: 4.53% used the internet as a source of information | Positive for frequent internet use, and for using the internet as a source of information | SRH mediated the relationship between internet use and community engagement |
| Yang et al 2020 [77] (USA) | Young and middle-aged: 26.8% excellent; 37.9% very good; 25.0% good; 8.2% fair; 2.1% poor<br>Older adults: 16.1% excellent; 33.3% very good; 33.4% good; 13.6% fair; 3.8% poor | Young and middle-aged: 85.6% daily; 6.1% use but not daily; 8.3% never<br>Older adults: 43.5% daily; 12.4% use but not daily; 44.1% never | Positive | Stronger among older adults than among young and middle-aged adults |

[a]NS: not significant.

[b]ADL: activities of daily living.

[c]N/A: not applicable.

[d]ICT: information and communication technology.

# Results

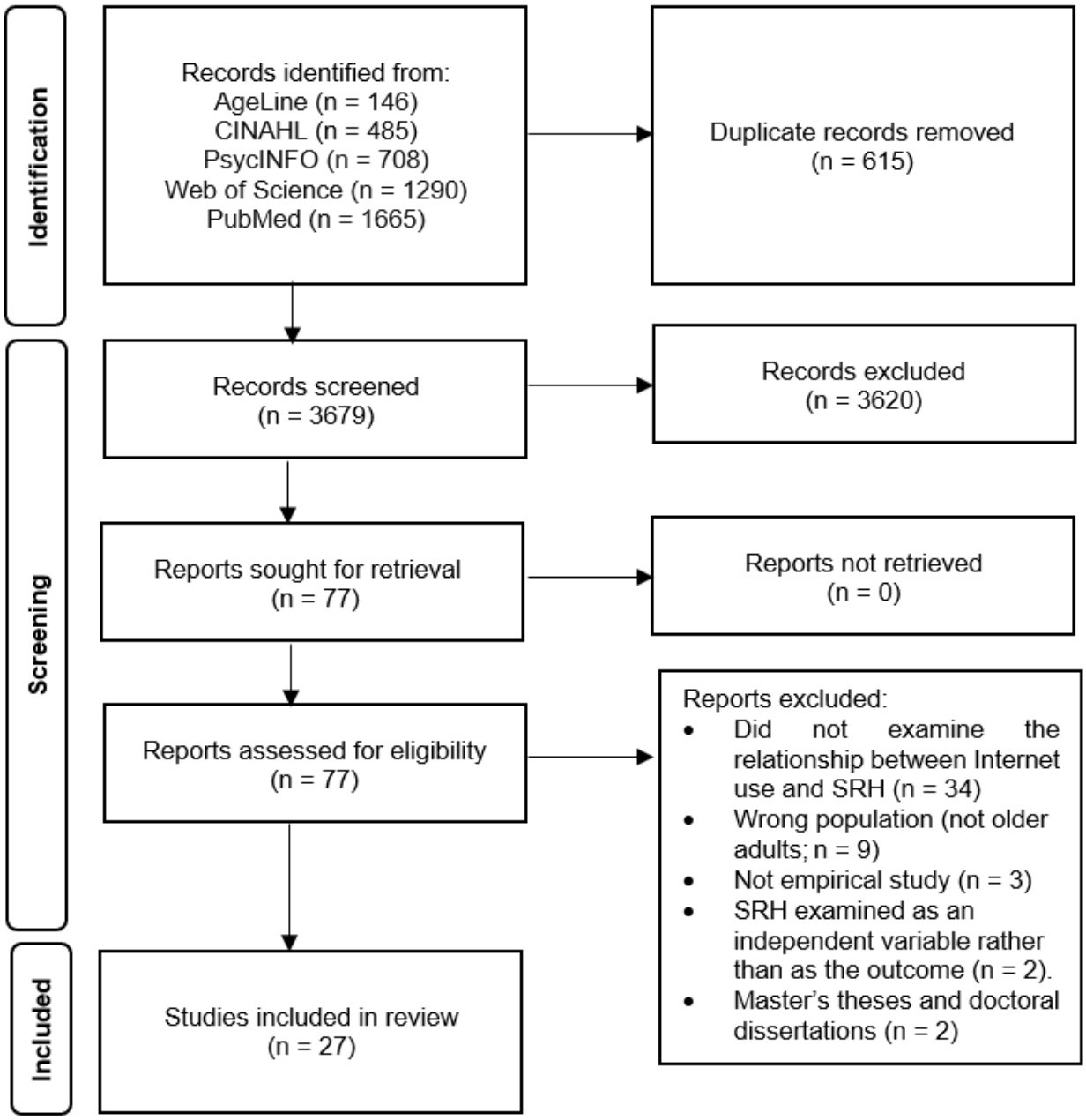
## Overview

The study selection process is shown in Figure 1. Searches of 5 databases identified 4294 records (PubMed, n=1665; Web of Science, n=1290; PsycINFO, n=708; CINAHL, n=485; AgeLine, n=146). After removing 615 duplicates, 3679 records were screened by title and abstract, and 77 full-text reports were assessed for eligibility. A total of 27 studies were included. The 50 excluded full-text reports were primarily excluded because they did not examine the internet use-SRH relationship (n=34), involved an ineligible population (n=9), were not empirical studies (n=3), treated SRH as an independent variable (n=2), or were theses/dissertations (n=2).



XSL•FO
RenderX

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram of study selection. SRH: self-rated health.




## Characteristics of Included Studies

### *Overview*

Table 1 provides an overview of the included studies and summarizes the following elements: study aims; publication year and country; study design and theoretical framework (where applicable); sample characteristics; and measures of SRH and internet use. Overall, the evidence base comprised predominantly cross-sectional, quantitative survey studies and included relatively limited use of explicit theoretical frameworks. Across studies, SRH was consistently assessed using a single-item measure (with varying response scales), whereas internet use was operationalized across multiple dimensions, including access/use (yes/no), frequency of use, and purpose-specific or domain-based measures.

### *Aims of the Studies*

Across nearly all included studies (26/27), internet use was examined as the independent variable and SRH as the outcome. In 1 study, SRH was examined as a mediator in the association between internet use and community participation [76].

### *Publication Years and Geographical Distribution*

Most included studies (n=22) were published between 2019 and 2024. Three studies were published between 2016 and 2018 (2016: n=2; 2018: n=1), and 2 studies were published in 2009. Studies were most frequently conducted in China (n=11) and the United States (n=7). Two studies were conducted in Korea, and 1 study each was conducted in Poland, Sweden, Norway, Spain, New Zealand, and Japan. One cross-national study analyzed data from 17 European countries and Israel [69].



XSL•FO
RenderX

### *Study Designs and Data Collation*

All included studies were quantitative and relied on survey or structured questionnaire data. **Notably,** no qualitative or mixed methods studies met the inclusion criteria. Several studies drew on the same underlying survey datasets: 3 studies used the Chinese General Social Survey (CGSS) [46,61,71], 5 studies used the China Health and Retirement Longitudinal Study (CHARLS) [49,59,60,75,76] (including 1 study that combined CHARLS with an additional dataset [59]), and 2 studies used the National Health and Aging Trends Study [55,58]. Two studies used longitudinal designs [66,71], while the remainder were cross-sectional. Specifically, Nakagomi et al [66] (2022) analyzed 3 waves of the Japan Gerontological Evaluation Study (2013, 2016, and 2019), and Wang et al [71] (2020) analyzed 2 waves of the CGSS (2012 and 2015).

### *Sample Characteristics*

Most included studies focused on community-dwelling populations. Although the review targeted older adults, 4 studies drew on datasets that included younger age groups in addition to older adults (eg, samples defined as ≥45, ≥50, or ≥35 years), while still reporting results specifically for older adults as a separate group [50,60,63,69]. In 1 study, older adults were explicitly compared with a younger group (18-64 years vs ≥65 years) [77]. One study focused on older cancer survivors [58], and 1 examined older military veterans [68].

Sample sizes varied widely. Four studies included fewer than 1000 participants (range 233-709) [48,52,53,67]. Most studies (n=18) used large samples of 1000-10,000 participants (range 1000-9434), including 1 study that analyzed 2 samples (n=598 and n=9434) [59]. Five studies used very large samples of >10,000 participants (range 13,474-82,014) [49,60,69,75,77].

### *Theoretical Framework*

Most included studies (21/27) did not report an explicit theoretical framework. Six studies referenced a theory or conceptual model: Chen et al [46] drew on activity theory; Wei et al [73] drew on social participation theory; Sims et al [67] referenced socioemotional selectivity theory and biological models of aging; Tavares [69] used a social ecological model; Millar et al [63] drew on Paasche-Orlow and Wolf's model and Gewald and Rockmann's model; and Xiaobing and Meng [76] drew on displacement theory. Overall, these findings indicate that the association between internet use and SRH has largely been examined in an atheoretical manner.

### *Measurements*

Overall, SRH was measured consistently across studies using a single-item indicator, although response scales varied. In contrast, internet use was assessed across 3 main dimensions: access/use (yes/no), frequency (ordinal scales), and purpose or domain of use (eg, social, instrumental, or health-related).

### *SRH Measurements*

All included studies assessed SRH using a single item, typically phrased as a global self-assessment of health. Response scales varied: 8 studies used dichotomous response options (eg, unhealthy/healthy; poor/good health; or satisfied/dissatisfied with health) [46,49,50,53,56-63,66]. Three studies used 3-point scales [68,72,75], 1 study used a 4-point scale [57], and 15 studies used 5-point scales [48,51,52,54,55,58-61,67,69,71,73,76,77].

### *Internet Use Measurements*

Measures of internet use varied substantially across studies. Nine studies measured internet use/access using a single binary indicator (eg, yes/no; user/nonuser; or use in the past 7 days) [46,49,52,53,56,61,62,69,75]. Nine studies assessed frequency of internet use using ordinal response scales ranging from 3-point to 7-point measures (including 3-point [77], 4-point [50,66,68], 5-point [71], and 7-point [51]). In addition, 3 studies used more than 1 internet-use measure [59,60,76]: Liu et al [60] included both a binary internet use indicator (yes/no) and a frequency measure (4-point scale); Xiaobing and Meng [76] included multiple internet-use indicators, including indicators of access/use, a measure of use frequency, and the internet as a primary source of information, and Li et al [59] used different frequency measures across samples (a 5-point scale in sample 1 and a 4-point scale in sample 2).

Several studies operationalized internet use in more specific ways, including socially oriented technology use [48], interpersonal communication and instrumental internet use [54], changes in internet use across multiple online domains [57], and internet use for health information seeking [63]. Finally, some studies assessed broader information and communication technology (ICT) use rather than internet use alone, including ICT indicators or indices and counts of devices or applications and related motivations [55,58,67]. Some studies also used platform- or device-specific measures, such as smartphone use or WeChat use [72,73].

## Main Results of Included Studies on Internet Use and SRH

### *Overview*

Table 2 summarizes the main findings of the included studies. It reports levels of SRH and internet use, the direction and significance of the association between internet use and SRH, and, where available, mediators and heterogeneity across subgroups. Overall, Table 2 indicates that most studies report a positive association between internet use and SRH. However, the magnitude and robustness of this association vary depending on the type of internet use examined and the characteristics of the study population.

### *Levels of SRH*

Across studies, SRH levels were reported either as mean scale scores or as proportions. In 12 studies reporting mean SRH scores, the average level of SRH was generally moderate to moderately high [48,52,54,55,57,58,60,61,67,71-73]. In studies reporting proportions, the share of participants reporting good to excellent health was often high (approximately 75% to the mid-80% range) [49,51,53,63,68], with an even higher proportion reported by Nakagomi et al [66] (87.8%). Lower proportions were also reported in some studies, including 53.47% reporting healthy status in the study by Lyu and Sun [62] and 39% reporting healthy status in the study by Chen et al [46].



XSL•FO
RenderX

### *Levels of Internet Use*

Overall, reported levels of internet use varied widely across studies, partly because internet use was operationalized in different ways (eg, binary use, frequency categories, or mean/index scores). Seven studies reported a relatively low prevalence of internet use, generally around 11%-24% [46,49,60-62,75,76]. In contrast, 4 studies reported substantially higher prevalence, with around 44%-57% of participants using the internet (or using a smart device) [52,56,69,72]. In studies reporting frequency-based measures, daily or near-daily use also varied considerably: 2 studies reported that roughly 25% of participants used the internet daily or almost every day [50,66], whereas 2 other studies reported higher daily-use rates including 48.7% [68] and 66.1% [51].

Several studies also reported purpose- or domain-specific levels of internet use. Social or communication-oriented use was reported as a low mean level of socially oriented technology use (mean 1.37, SD 1.42) [48], moderate to high engagement in communication technology in some samples (eg, emailing/texting most days, 65.5%) [55], and lower engagement in others (eg, emailing/texting some days, 15.7%, and most days, 26.8%) [58]. In 1 study assessing online purposes, interpersonal communication use was reported (mean 2.24, SD 0.74), alongside instrumental use (mean 3.22, SD 2.22) [54], and another study reported moderate to moderately high levels of change across multiple online domains (approximate means 2.87-3.26 across domains) [57]. Health-related internet use also varied: in 1 study, specific health-related online activities ranged from 11.8% to 36.2% [55], whereas another study reported 69.96% using the internet for health information seeking [63]. Finally, platform- or device-specific use was generally low for WeChat (5%) [73] and higher for smartphone use (57%) [72].

### *The Association Between Internet Use and SRH*

The majority of included studies (24/27) reported a positive association between internet use and SRH, indicating that older adults who used the internet tended to report better SRH than nonusers [46,48,49,51-62,66-69,72,73,75-77]. Consistent with this pattern, studies assessing frequency-based measures generally suggested that more frequent internet use was associated with better SRH [54,57,59,66-68,76,77]. Purpose-specific measures showed positive associations in several contexts, including socially oriented technology use and communication-related use [48,54,57,58], and instrumental use, such as accessing information and services, in some studies [54,57]. One study also reported that combined ICT use was associated with better SRH [55].

At the same time, findings were not uniform. One study reported a modest association that was significant primarily for near-daily internet use [66], and another found that communication technology was positively associated with SRH only when used most days [58]. One study reported that the contribution of internet use to SRH was minor compared with health-related factors, such as chronic or psychiatric conditions [51], and another found that the association was attenuated after accounting for social class [53]. Two studies reported no significant association between internet use and SRH [63,71], and 1 study reported a negative association [50]. In addition, nonsignificant results were reported for specific types of health-related internet use in some studies [58,63].

### *Mediating Factors in the Relationship Between Internet Use and SRH*

Six of the included studies (6/27) examined mediating pathways linking internet use with SRH. Several of these studies highlighted social pathways: cultural engagement [46], social support from relatives and friends [61], and social engagement [60] were each reported as mediators of the association between internet use and SRH. Loneliness was also identified as a mediator in 1 study, with socially oriented technology use associated with lower loneliness, which in turn was linked to better SRH [48]. In addition, 1 study reported that using technology to learn new information and skills mediated the association between ICT use and SRH, whereas using technology to connect with family and friends did not show a mediating effect [67]. Finally, 1 study identified performance expectations and individualized needs as mediators of the association between smart device use and SRH [72].

### *Heterogeneity Across Subgroups in the Association Between Internet Use and SRH*

Studies examining subgroup differences in the association between internet use and SRH reported heterogeneous findings. With respect to age, 1 study comparing younger and older adults reported a stronger association between smartphone use and SRH among older adults than among younger and middle-aged adults [77]. In addition, age-stratified analyses in 1 study of internet use domains showed that social networking and information sharing and daily services were positively associated with SRH in both the young-old (aged 65-74 years) and old-old (aged ≥75 years) groups, whereas social participation services and search, email, and content services were positively associated with SRH only in the old-old group [57]. Other studies also suggested that the benefits of internet and social media use may be more pronounced among younger-old adults (eg, aged <70 years) [61,73].

Regarding place of residence, findings were mixed: 1 study reported a stronger benefit of internet use among rural older adults [61], whereas another reported a stronger association between smartphone use and SRH among urban older adults [72]. A third study found no differences between urban and rural settings [75]. For gender, 1 study reported a stronger benefit of internet use among older men than older women [61], whereas another study found no gender differences in the association [73]. Evidence of heterogeneity was also reported for other population characteristics. Internet access appeared more beneficial for vulnerable groups (eg, those with difficulties in activities of daily living) and for individuals with lower social participation or without a spouse [49]. In contrast, 1 study reported that internet use was not beneficial among older adults with low levels of health [69], while also indicating a stronger association between internet use and SRH in countries with less developed eHealth contexts [69]. In 1 study, cognitive ability was examined as a moderator, but no significant moderating effect was found [71]. Finally, 1 study reported that the contribution of internet use to SRH was greater among older adults with children than among those without children [60].



XSL•FO
RenderX

## Discussion

### Principal Findings

The purpose of this scoping review was to map and characterize the existing literature on internet use and SRH among older adults, including how this relationship has been studied, what patterns of association have been reported, and which gaps remain in the evidence. Overall, the review included 27 studies, most of which (24/27) reported a statistically significant positive association between internet use and SRH among older adults. Several studies further suggested that socially oriented uses of the internet, including communication and social participation, were positively associated with SRH [48,57,58]. Evidence on potential mechanisms was more limited; however, the available findings indicate that social factors, including social support and social participation, may help explain the observed association between internet use and SRH [46,60-62]. In addition, some studies suggested that certain vulnerable groups, such as individuals with difficulties in activities of daily living, lower social participation, or without a partner, may derive greater benefit from internet use, although subgroup findings were mixed overall [49]. These findings can be interpreted in light of social participation theory, which posits that active engagement in social activities, whether in-person or virtual, can enhance well-being and contribute to better perceived health in later life. Similarly, social capital theory highlights the importance of social networks and the resources they provide, including emotional and instrumental support [78]; these resources may be expanded or sustained through internet use and, in turn, relate to SRH. Beyond these perspectives, as a possible theoretical interpretation, the findings may also be viewed through the lens of socioemotional selectivity theory [79], which suggests that as individuals age, they increasingly prioritize emotionally meaningful goals and relationships. From this perspective, socially oriented internet use, including communication with family and friends or participation in online social activities, may be particularly relevant to older adults' SRH because it supports emotionally salient connections that are central to well-being in later life.

At first glance, these findings suggest that interventions supporting older adults in acquiring internet skills may be relevant for supporting SRH. However, a closer examination highlights persistent digital gaps. Several studies in this review reported relatively low levels of internet use, particularly in studies that included a binary indicator of internet use (eg, Chen et al [46], Liu et al [60], Lyu and Sun [62], and Xiaobing and Meng [76]). Although internet adoption among older adults has increased over the past decade [33-35], evidence continues to point to a digital divide [80,81]. Digital divide frameworks conceptualize digital inclusion as a cumulative process that involves motivation to use digital technologies, physical access, and digital skills [82]. Consistent with this perspective, the divide reflects not only age-related barriers (ie, "gray" gaps), including negative attitudes toward technology, limited digital skills, lack of interest, and experiences of ageism [33,83,84], but also broader sociodemographic and structural disparities, including lower socioeconomic resources, lower education, poorer health, and disability [34-46,48-63,66-69,71-73,75-89]. As a result, it is important that interventions address access conditions, including infrastructure, devices, and affordability, as well as guidance and support that strengthen digital skills and enable safe and meaningful use. Such a combined approach may help reduce inequalities in opportunities for social connection, access to information and services, and potential health-related benefits.

Indeed, most studies in the review reported a positive association between internet use and SRH; however, 1 study reported a negative association [50], and 2 studies reported no significant relationship [63,71]. Duplaga [50] found that more frequent internet users were more likely to rate their health as poor, possibly reflecting a more critical self-appraisal associated with exposure to health-related content. It is therefore important to acknowledge that internet use may also entail potential risks for older adults. One concern is exposure to online health misinformation, which can lead to confusion, mistrust in health professionals, and potentially harmful health behaviors [90,91]. Additional risks include digital fatigue and information overload, which may contribute to cognitive strain and mental exhaustion [91-93], as well as cybersecurity threats, such as scams, fraud, and privacy violations, which can increase stress and reduce trust in digital technologies [94,95]. Together, these considerations highlight the need for digital literacy programs and supportive interventions to help ensure that internet use promotes, rather than undermines, older adults' health and well-being. Finally, the absence of a positive association between internet use and SRH in some studies may reflect background and contextual factors, such as socioeconomic resources (eg, income and education), cognitive ability, and the inherently subjective and multidimensional nature of SRH, which may be less responsive to behavioral influences such as internet use [63,71].

Despite the importance of the results of the included studies, we identified several research gaps that limit our understanding of the relationship between internet use and SRH. First, it is important to note that a significant portion of the studies included in this review (25/27) were based on a cross-sectional design, a common methodological approach that, by its nature, limits the ability to infer causal relationships between variables because it captures phenomena at a single point in time [96]. This methodological limitation requires careful consideration when interpreting the findings and generalizing them to dynamic or process-based relationships. Nevertheless, the included studies met most of the key criteria for assessing risk of bias in cross-sectional research, as outlined in the methodological review by Kelly et al [97], including selection (sample representativeness), exposure (measurement validity), outcome (measurement reliability), confounding (control of variables), missingness (handling missing data), selective reporting (complete results disclosure), and conflict of interest (financial or professional influence). The overall methodological quality strengthens the degree of confidence that can be placed in the review's findings, even if conclusions must be drawn with caution. In light of the methodological challenges inherent to this design, there is a clear need to promote longitudinal and interventional studies that allow for a deeper understanding of dynamic relationships and causal pathways over time. Second,



XSL•FO
RenderX

most of the studies did not apply a theoretical framework or model. Only 6 of the 27 studies referred to a theory, such as activity theory, social participation theory, or the social ecological model. The limited use of theoretical frameworks may reflect the lack of models that are specifically tailored to understanding the relationship between internet use and SRH in older adults. Without a guiding framework, it becomes more difficult to interpret the processes underlying this relationship, including possible mediators and moderators. Therefore, future research should aim to develop or adopt theoretical models that can offer clearer conceptual guidance for investigating this association. Third, regarding internet use, the majority of the studies focused primarily on examining use or frequency of use without reference to the purpose of internet use. Therefore, further studies should focus on understanding how the different purposes for which the internet is used may affect older adults' SRH. Finally, this review included only studies from industrialized countries, highlighting the need for future research in developing countries and low- and middle-income countries. In these contexts, internet access is often more limited, and older adults may encounter unique barriers such as lower levels of digital literacy, limited technological infrastructure, and different social and health care systems [98]. In addition, sociocultural factors such as norms surrounding aging, family roles, and attitudes toward technology may also influence patterns of internet use and its impact on SRH [99,100]. Therefore, findings from low- and middle-income countries may differ substantially from those in high-income countries and warrant focused, context-sensitive investigation.

### Limitations

This scoping review has several limitations. Although the search strategy followed established guidelines and covered 5 major electronic databases, supplemented by Google Scholar searches and reference list screening, relevant studies published in sources not indexed in these databases may have been missed. In addition, the review was restricted to English-language publications; therefore, relevant studies published in other languages may have been excluded. Finally, the review included only peer-reviewed journal articles and did not incorporate gray literature, despite evidence that gray literature can provide valuable insights in systematic and scoping reviews [101].

### Conclusions and Implications for Practice and Policy

This scoping review mapped 27 studies examining the association between internet use and SRH among older adults. Most studies reported a positive association between internet use and better SRH, although the magnitude and consistency of this relationship varied depending on how internet use was measured and on the characteristics of specific subgroups. Overall, the findings suggest that the relationship between internet use and SRH depends substantially on the purpose of use, with socially oriented uses such as communication and social participation appearing particularly relevant and potentially operating through mediating social pathways, including greater perceived support and reduced loneliness. However, because the evidence base is dominated by cross-sectional studies, longitudinal and intervention research is needed to strengthen causal inference, strengthen and extend the evidence on mediating mechanisms, which have been examined in a relatively small number of studies and mainly through social factors, and on moderating mechanisms, and to clarify for whom and under what conditions internet use contributes to SRH.

From a practical standpoint, the findings support advancing digital inclusion as part of healthy aging policy by providing age-tailored training that promotes safe and meaningful use, especially uses that facilitate social connection, access to information and services, and community participation. Digital support should be integrated into health and social care through community-based initiatives and digital navigators who can provide hands-on guidance and ongoing assistance. In addition, improving digital inclusion may also benefit health and social care systems by facilitating timely access to services and support and potentially contributing to greater system efficiency [102]. This may also help reduce health disparities among older adults [102]. At the same time, maintaining nondigital options across service points and information channels is important to prevent exclusion of those who do not use the internet or struggle to do so. These recommendations align with framing digital inclusion as a social determinant of health rather than merely a technological issue [102]. Finally, there remains a notable gap in evidence from low- and middle-income countries, where sociocultural contexts and system-level conditions may shape patterns of internet use and its health implications [99,100]. In these settings, and among socioeconomically disadvantaged groups within high-income countries, digital inclusion efforts may also need to address structural barriers such as affordability, access to connectivity, and the availability of accessible online services tailored to older adults.

## Data Availablilty

The datasets used during this study are available from the corresponding author on reasonable request.

## Funding

The author received no financial support for the research, authorship, and/or publication of this study.

## Authors' Contributions

HA led the conceptualization, methodology, investigation, data curation, formal analysis, visualization, project administration, and the writing of the original draft, as well as the review and editing. FA-R led validation and contributed in a supporting role



XSL•FO
RenderX

to methodology, data curation, formal analysis, and the review and editing of the manuscript. Both authors approved the final manuscript.

The authors declare the use of generative artificial intelligence (GAI) in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GAI tools under full human supervision: proofreading and editing.

The GAI tool used was ChatGPT 5.2 (OpenAI). Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Declaration submitted by HA.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR checklist.
[PDF File (Adobe PDF File), 176 KB - ijmr_v15i1e76930_app1.pdf ]

Multimedia Appendix 2
PRISMA-S checklist.
[PDF File (Adobe PDF File), 261 KB - ijmr_v15i1e76930_app2.pdf ]

Multimedia Appendix 3
Search strategies.
[PDF File (Adobe PDF File), 103 KB - ijmr_v15i1e76930_app3.pdf ]

## References


1. Fayers PM, Sprangers MA. Understanding self-rated health. Lancet 2002;359(9302):187-188. [doi: 10.1016/S0140-6736(02)07466-4] [Medline: 11812551]
2. Murata C, Kondo T, Tamakoshi K, Yatsuya H, Toyoshima H. Determinants of self-rated health: could health status explain the association between self-rated health and mortality? Arch Gerontol Geriatr 2006;43(3):369-380. [doi: 10.1016/j.archger.2006.01.002] [Medline: 16530864]
3. Jylhä M. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. Soc Sci Med 2009;69(3):307-316. [doi: 10.1016/j.socscimed.2009.05.013] [Medline: 19520474]
4. Bombak AE. Self-rated health and public health: a critical perspective. Front Public Health 2013;1:15 [FREE Full text] [doi: 10.3389/fpubh.2013.00015] [Medline: 24350184]
5. Ocampo JM. Self-rated health: importance of use in elderly adults. Colomb Med 2010;41(3):275-289. [doi: 10.25100/cm.v41i3.715]
6. Lin YH, Chen HC, Hsu NW, Chou P. Validation of global self-rated health and happiness measures among older people in the Yilan study, Taiwan. Front Public Health 2020;8:346 [FREE Full text] [doi: 10.3389/fpubh.2020.00346] [Medline: 32850586]
7. Bamia C, Orfanos P, Juerges H, Schöttker B, Brenner H, Lorbeer R, et al. Self-rated health and all-cause and cause-specific mortality of older adults: individual data meta-analysis of prospective cohort studies in the CHANCES consortium. Maturitas 2017;103:37-44. [doi: 10.1016/j.maturitas.2017.06.023] [Medline: 28778331]
8. Li CL, Chang HY, Wang HH, Bai YB. Diabetes, functional ability, and self-rated health independently predict hospital admission within one year among older adults: a population based cohort study. Arch Gerontol Geriatr 2011;52(2):147-152. [doi: 10.1016/j.archger.2010.03.004] [Medline: 20338646]
9. Nägga K, Dong HJ, Marcusson J, Skoglund SO, Wressle E. Health-related factors associated with hospitalization for old people: comparisons of elderly aged 85 in a population cohort study. Arch Gerontol Geriatr 2012;54(2):391-397. [doi: 10.1016/j.archger.2011.04.023] [Medline: 21640394]
10. Wuorela M, Lavonius S, Salminen M, Vahlberg T, Viitanen M, Viikari L. Self-rated health and objective health status as predictors of all-cause mortality among older people: a prospective study with a 5-, 10-, and 27-year follow-up. BMC Geriatr 2020;20(1):120 [FREE Full text] [doi: 10.1186/s12877-020-01516-9] [Medline: 32228464]
11. Zadworna M. Pathways to healthy aging - exploring the determinants of self-rated health in older adults. Acta Psychol (Amst) 2022;228:103651 [FREE Full text] [doi: 10.1016/j.actpsy.2022.103651] [Medline: 35785683]
12. Arnadottir SA, Gunnarsdottir ED, Stenlund H, Lundin-Olsson L. Determinants of self-rated health in old age: a population-based, cross-sectional study using the International Classification of Functioning. BMC Public Health 2011;11:670 [FREE Full text] [doi: 10.1186/1471-2458-11-670] [Medline: 21867517]




XSL•FO
RenderX

13. Badri MA, Yang G, Al Khaili M, Al Bahar M, Al Rashdi A, Al Hyas L. Hierarchical regression of wellbeing and self-rated health among older adults in Abu Dhabi. Int J Environ Res Public Health 2021;18(15):8006 [FREE Full text] [doi: 10.3390/ijerph18158006] [Medline: 34360297]
14. Josefsson K, Andersson M, Erikstedt A. Older adults' self-rated health and differences by age and gender: a quantitative study. Healthy Aging Res 2016;5(5). [doi: 10.12715/har.2016.5.1]
15. Meng Q, Xie Z, Zhang T. A single-item self-rated health measure correlates with objective health status in the elderly: a survey in suburban Beijing. Front Public Health 2014;2:27 [FREE Full text] [doi: 10.3389/fpubh.2014.00027] [Medline: 24783187]
16. Schneider G, Driesch G, Kruse A, Wachter M, Nehen HG, Heuft G. What influences self-perception of health in the elderly? The role of objective health condition, subjective well-being and sense of coherence. Arch Gerontol Geriatr 2004;39(3):227-237. [doi: 10.1016/j.archger.2004.03.005] [Medline: 15381341]
17. Stanojevic Jerkovic O, Sauliune S, Šumskas L, Birt C, Kersnik J. Determinants of self-rated health in elderly populations in urban areas in Slovenia, Lithuania and UK: findings of the EURO-URHIS 2 survey. Eur J Public Health 2017;27(suppl_2):74-79. [doi: 10.1093/eurpub/ckv097] [Medline: 26163468]
18. Aggarwal B, Xiong Q, Schroeder-Butterfill E. Impact of the use of the internet on quality of life in older adults: review of literature. Prim Health Care Res Dev 2020;21:e55 [FREE Full text] [doi: 10.1017/S1463423620000584] [Medline: 33263273]
19. Cotten SR, Anderson WA, McCullough BM. Impact of internet use on loneliness and contact with others among older adults: cross-sectional analysis. J Med Internet Res 2013;15(2):e39 [FREE Full text] [doi: 10.2196/jmir.2306] [Medline: 23448864]
20. Heo J, Chun S, Lee S, Lee KH, Kim J. Internet use and well-being in older adults. Cyberpsychol Behav Soc Netw 2015;18(5):268-272. [doi: 10.1089/cyber.2014.0549] [Medline: 25919967]
21. Hülür G, Macdonald B. Rethinking social relationships in old age: digitalization and the social lives of older adults. Am Psychol 2020;75(4):554-566. [doi: 10.1037/amp0000604] [Medline: 32378949]
22. Yang C, Lai DWL, Sun Y, Ma C, Chau AKC. Mobile application use and loneliness among older adults in the digital age: insights from a survey in Hong Kong during the COVID-19 pandemic. Int J Environ Res Public Health 2022;19(13):7656 [FREE Full text] [doi: 10.3390/ijerph19137656] [Medline: 35805316]
23. Zhou R, Fong PSW, Tan P. Internet use and its impact on engagement in leisure activities in China. PLoS One 2014;9(2):e89598 [FREE Full text] [doi: 10.1371/journal.pone.0089598] [Medline: 24586902]
24. Estacio EV, Whittle R, Protheroe J. The digital divide: examining socio-demographic factors associated with health literacy, access and use of internet to seek health information. J Health Psychol 2019;24(12):1668-1675. [doi: 10.1177/1359105317695429] [Medline: 28810415]
25. Nordin S, Sturge J, Ayoub M, Jones A, McKee K, Dahlberg L, et al. The role of information and communication technology (ICT) for older adults' decision-making related to health, and health and social care services in daily life-a scoping review. Int J Environ Res Public Health 2021;19(1):151 [FREE Full text] [doi: 10.3390/ijerph19010151] [Medline: 35010408]
26. Zhao YC, Zhao M, Song S. Online health information seeking behaviors among older adults: systematic scoping review. J Med Internet Res 2022;24(2):e34790 [FREE Full text] [doi: 10.2196/34790] [Medline: 35171099]
27. Cao Y, Sun L, Luo Y, Li M, Tian W. Relationship between typology of internet users and social networks in Chinese older adults. J Appl Gerontol 2023;42(5):992-1002. [doi: 10.1177/07334648231151283] [Medline: 36639849]
28. Carlo S, Vergani M. Risk and benefit perceptions: resistance, adoption and uses of ICT among the Italian elderly. 2016 Presented at: Second International Conference, ITAP 2016, held as part of HCI International 2016; 2016 July 17–22; Toronto, ON, Canada p. 155-166. [doi: 10.1007/978-3-319-39943-0_15]
29. Tang D, Jin Y, Zhang K, Wang D. Internet use, social networks, and loneliness among the older population in China. Front Psychol 2022;13:895141 [FREE Full text] [doi: 10.3389/fpsyg.2022.895141] [Medline: 35645921]
30. Wang D, Liu X, Chen K, Gu C, Zhao H, Zhang Y, et al. Risks and protection: a qualitative study on the factors for internet addiction among elderly residents in Southwest China communities. BMC Public Health 2024;24(1):531 [FREE Full text] [doi: 10.1186/s12889-024-17980-6] [Medline: 38378524]
31. Alagood J, Prybutok G, Prybutok VR. Navigating privacy and data safety: the implications of increased online activity among older adults post-COVID-19 induced isolation. Information 2023;14(6):346. [doi: 10.3390/info14060346]
32. Tao J, Shuijing H. The elderly and the big data how older adults deal with digital privacy. : IEEE; 2016 Presented at: International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS); 2016 December 17-18; Changsha, China p. 285-288. [doi: 10.1109/icitbs.2016.35]
33. Charness N, Boot WR. A grand challenge for psychology: reducing the age-related digital divide. Curr Dir Psychol Sci 2022;31(2):187-193 [FREE Full text] [doi: 10.1177/09637214211068144] [Medline: 35754678]
34. Hunsaker A, Hargittai E. A review of internet use among older adults. New Media Soc 2018;20(10):3937-3954. [doi: 10.1177/1461444818787348]
35. Yoon H, Jang Y, Kim S, Speasmaker A, Nam I. Trends in internet use among older adults in the United States, 2011-2016. J Appl Gerontol 2021;40(5):466-470. [doi: 10.1177/0733464820908427] [Medline: 32131670]
36. Riazul Islam SM, Kwak D, Kabir MH, Hossain M, Kwak KS. The Internet of Things for health care: a comprehensive survey. IEEE Access 2015;3:678-708. [doi: 10.1109/access.2015.2437951]




XSL•FO
RenderX

37. Kashani MH, Madanipour M, Nikravan M, Asghari P, Mahdipour E. A systematic review of IoT in healthcare: applications, techniques, and trends. J Netw Comput Appl 2021;192:103164. [doi: 10.1016/j.jnca.2021.103164]
38. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York: Guilford Publications; 2017.
39. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
40. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, PRISMA-S Group. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev 2021;10(1):39 [FREE Full text] [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
41. Falagas ME, Paliogianni PM, Kontogiannis DS, Ragias D, Johnson E. Google Scholar as a resource for systematic reviews in clinical medicine. J Eval Clin Pract 2025;31(5):e70206. [doi: 10.1111/jep.70206] [Medline: 40652554]
42. Huang J, Zhang Y, Xv M, Sun L, Wang M. Association between oral health status and frailty in older adults: a systematic review and meta-analysis. Front Public Health 2025;13:1514623 [FREE Full text] [doi: 10.3389/fpubh.2025.1514623] [Medline: 40231179]
43. Landon SN, Padikkala J, Horwitz LI. Identifying drivers of health care value: a scoping review of the literature. BMC Health Serv Res 2022;22(1):845 [FREE Full text] [doi: 10.1186/s12913-022-08225-6] [Medline: 35773663]
44. Imes CC, Barthel NJ, Chasens ER, Dunbar-Jacob J, Engberg SJ, Feeley CA, et al. Shift work organization on nurse injuries: a scoping review. Int J Nurs Stud 2023;138:104395. [doi: 10.1016/j.ijnurstu.2022.104395] [Medline: 36481596]
45. Pollock D, Peters MDJ, Khalil H, McInerney P, Alexander L, Tricco AC, et al. Recommendations for the extraction, analysis, and presentation of results in scoping reviews. JBI Evid Synth 2023;21(3):520-532. [doi: 10.11124/JBIES-22-00123] [Medline: 36081365]
46. Chen WC, Yang L, Wang XY. Internet use, cultural engagement, and multi-dimensional health of older adults: a cross-sectional study in China. Front Public Health 2022;10:887840 [FREE Full text] [doi: 10.3389/fpubh.2022.887840] [Medline: 35692319]
47. Fernández-Ballesteros R, Dolores Zamarrón M, Angel Ruíz M. The contribution of socio-demographic and psychosocial factors to life satisfaction. Ageing Soc 2001;21(1):25-43. [doi: 10.1017/S0144686X01008078]
48. Chopik WJ. The benefits of social technology use among older adults are mediated by reduced loneliness. Cyberpsychol Behav Soc Netw 2016;19(9):551-556 [FREE Full text] [doi: 10.1089/cyber.2016.0151] [Medline: 27541746]
49. Ding X, Yuan L, Zhou Y. Internet access and older adults' health: evidence from China. China Econ Rev 2023;82:102047. [doi: 10.1016/j.chieco.2023.102047]
50. Duplaga M. The association between Internet use and health-related outcomes in older adults and the elderly: a cross-sectional study. BMC Med Inform Decis Mak 2021;21(1):150 [FREE Full text] [doi: 10.1186/s12911-021-01500-2] [Medline: 33957909]
51. Falk Erhag H, Ahlner F, Rydberg Sterner T, Skoog I, Bergström A. Internet use and self-rated health among Swedish 70-year-olds: a cross-sectional study. BMC Geriatr 2019;19(1):365. [doi: 10.1186/s12877-019-1392-8] [Medline: 31870313]
52. Fjell A, Cronfalk Seiger B, Hermann M, Rongve A, Aßmus J, Kvinge L, Seiger, et al. Factors associated with self-rated health in a Norwegian population of older people participating in a preventive home visit program: a cross-sectional study. BMC Geriatr 2020;20(1):323 [FREE Full text] [doi: 10.1186/s12877-020-01733-2] [Medline: 32887555]
53. Gracia E, Herrero J. Internet use and self-rated health among older people: a national survey. J Med Internet Res 2009;11(4):e49 [FREE Full text] [doi: 10.2196/jmir.1311] [Medline: 19955041]
54. Jeon GS, Choi K. Purposes of internet use and its impacts on physical and psychological health of Korean older adults. Healthcare (Basel) 2024;12(2):244 [FREE Full text] [doi: 10.3390/healthcare12020244] [Medline: 38255131]
55. Kim J, Lee HY, Won CR, Barr T, Merighi JR. Older adults' technology use and its association with health and depressive symptoms: findings from the 2011 National Health and Aging Trends Study. Nurs Outlook 2020;68(5):560-572 [FREE Full text] [doi: 10.1016/j.outlook.2020.05.001] [Medline: 32527596]
56. Koopman-Boyden PG, Reid SL. Internet/e-mail usage and well-being among 65–84 year olds in New Zealand: policy implications. Educ Gerontol 2009;35(11):990-1007. [doi: 10.1080/03601270902917745]
57. Lee J, Jang SN. Have changes in internet use during the COVID-19 pandemic affected older adults' self-rated health? A cross-sectional study of young-old and old-old populations in Korea. Geriatr Nurs 2022;48:145-149 [FREE Full text] [doi: 10.1016/j.gerinurse.2022.09.012] [Medline: 36219933]
58. Lee HY, Kim J, Sharratt M. Technology use and its association with health and depressive symptoms in older cancer survivors. Qual Life Res 2018;27(2):467-477. [doi: 10.1007/s11136-017-1734-y] [Medline: 29128998]
59. Li C, Mo W, Wei Q. The role of internet use and offline social engagement in the health of Chinese older adults: evidence from nationally representative samples. Healthcare (Basel) 2023;11(5):653 [FREE Full text] [doi: 10.3390/healthcare11050653] [Medline: 36900658]
60. Liu W, Li W, Mou J. Does internet usage make middle-aged and older adults feel healthier? Mediating role of social engagement. Ind Manag Data Syst 2023;124(1):1-28. [doi: 10.1108/imds-04-2023-0236]




XSL•FO
RenderX

61. Liu N, He Y, Li Z. The relationship between internet use and self-rated health among older adults in China: the mediating role of social support. Int J Environ Res Public Health 2022;19(22):14785 [FREE Full text] [doi: 10.3390/ijerph192214785] [Medline: 36429504]
62. Lyu S, Sun J. Internet use and self-rated health among Chinese older adults: the mediating role of social capital. Geriatr Gerontol Int 2021;21(1):34-38. [doi: 10.1111/ggi.14090] [Medline: 33280230]
63. Millar RJ, Sahoo S, Yamashita T, Cummins P. Problem solving in technology-rich environments and self-rated health among adults in the U.S.: an analysis of the Program for the International Assessment of Adult Competencies. J Appl Gerontol 2020;39(8):889-897 [FREE Full text] [doi: 10.1177/0733464819829663] [Medline: 30762453]
64. Paasche-Orlow M, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav 2007;31:S19-S26. [doi: 10.5993/AJHB.31.s1.4]
65. Gewald H, Rockmann R. Mature adults' use of digital health services the role of prior computer experience on eHealth adoption. 2016 Presented at: CONF-IRM 2016 Proceedings. 1; 2016 May 18-20; Cape Town URL: https://aisel.aisnet.org/confirm2016/1/
66. Nakagomi A, Shiba K, Kawachi I, Ide K, Nagamine Y, Kondo N, et al. Internet use and subsequent health and well-being in older adults: an outcome-wide analysis. Comput Human Behav 2022;130:107156. [doi: 10.1016/j.chb.2021.107156]
67. Sims T, Reed AE, Carr DC. Information and communication technology use is related to higher well-being among the oldest-old. J Gerontol B Psychol Sci Soc Sci 2017;72(5):761-770. [doi: 10.1093/geronb/gbw130] [Medline: 27702839]
68. Swed O, Sheehan CM, Butler JS. The digital divide and veterans' health: differences in self-reported health by internet usage. Armed Forces Soc 2018;46(2):238-258. [doi: 10.1177/0095327x18809069]
69. Tavares AI. Self-assessed health among older people in Europe and internet use. Int J Med Inform 2020;141:104240. [doi: 10.1016/j.ijmedinf.2020.104240] [Medline: 32739610]
70. Policies and strategies to promote social equity in health. Institute for Futures Studies. 1991. URL: https://core.ac.uk/download/pdf/6472456.pdf [accessed 2026-02-02]
71. Wang J, Liang C, Li K. Impact of internet use on elderly health: empirical study based on chinese general social survey (CGSS) data. Healthcare (Basel) 2020;8(4):482 [FREE Full text] [doi: 10.3390/healthcare8040482] [Medline: 33198418]
72. Wei Y, Guo X. Impact of smart device use on objective and subjective health of older adults: findings from four provinces in China. Front Public Health 2023;11:1118207 [FREE Full text] [doi: 10.3389/fpubh.2023.1118207] [Medline: 37546303]
73. Wei N, Sun D, Huang W. Effects of WeChat use on the subjective health of older adults. Front Psychol 2022;13:919889 [FREE Full text] [doi: 10.3389/fpsyg.2022.919889] [Medline: 36172240]
74. Lian WM, Gan G, Pin C, Wee S, Ye H. Correlates of leisure-time physical activity in an elderly population in Singapore. Am J Public Health 1999;89(10):1578-1580. [doi: 10.2105/AJPH.89.10.1578]
75. Wen W, Zhang Y, Shi W, Li J. Association between internet use and physical health, mental health, and subjective health in middle-aged and older adults: nationally representative cross-sectional survey in China. J Med Internet Res 2023;25:e40956 [FREE Full text] [doi: 10.2196/40956] [Medline: 36943368]
76. Xiaobing H, Meng C. The impact of internet use on community participation of older adults: evidence from China. Sage Open 2022;12(2):21582440221097387. [doi: 10.1177/21582440221097387]
77. Yang Y, Li S, Zhang K, Xiang X, Li Z, Ahn S, et al. How the daily smartphone is associated with daily travel, physical activity, and self-perceived health: evidence from 2017 national household travel survey. J Aging Phys Act 2020;28(5):740-748. [doi: 10.1123/japa.2019-0222] [Medline: 32434149]
78. Li PQ, Sun GY, Zhao XX, Hu ZX, Gan KP. Internet use and self-rated health among Chinese older adults: the role of social engagement and social support. Front Public Health 2024;12:1504940 [FREE Full text] [doi: 10.3389/fpubh.2024.1504940] [Medline: 39975695]
79. Carstensen LL, Isaacowitz DM, Charles ST. Taking time seriously. A theory of socioemotional selectivity. Am Psychol 1999;54(3):165-181. [doi: 10.1037//0003-066x.54.3.165] [Medline: 10199217]
80. Blažič BJ, Blažič AJ. Overcoming the digital divide with a modern approach to learning digital skills for the elderly adults. Educ Inf Technol 2019;25(1):259-279. [doi: 10.1007/s10639-019-09961-9]
81. Liu L, Wu F, Tong H, Hao C, Xie T. The digital divide and active aging in China. Int J Environ Res Public Health 2021;18(23):12675 [FREE Full text] [doi: 10.3390/ijerph182312675] [Medline: 34886400]
82. van Dijk JA. Digital divide research, achievements and shortcomings. Poetics 2006;34(4-5):221-235. [doi: 10.1016/j.poetic.2006.05.004]
83. Choi EY, Kim Y, Chipalo E, Lee HY. Does perceived ageism widen the digital divide? And does it vary by gender? Gerontologist 2020;60(7):1213-1223 [FREE Full text] [doi: 10.1093/geront/gnaa066] [Medline: 32539099]
84. Alonso Gonzalez D, D'Antonio Maceiras S, Diaz Catalan C, Sadaba Rodriguez I. Types of older adults ICT users and the grey divide: attitudes matter. RCIS 2021;74:120-137. [doi: 10.33788/rcis74.8]
85. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme version. CloudFront. 2006. URL: https://d1wqtxts1xzle7.cloudfront.net/39246301/02e7e5231e8f3a6183000000-libre.pdf?1445068016=&response-content-disposition=inline%3B+filename%3DGuidance_on_the_conduct_of_narrative_syn.pdf&Expires=1770911173&Signature=Hj-IBcxhdgv9fT9LPztL0vm3Rhe2nfFDkvmb [accessed 2026-02-02]




XSL•FO
RenderX

86. Chang J, McAllister C, McCaslin R. Correlates of, and barriers to, Internet use among older adults. J Gerontol Soc Work 2015;58(1):66-85. [doi: 10.1080/01634372.2014.913754] [Medline: 24941050]
87. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
88. Hargittai E, Piper AM, Morris MR. From internet access to internet skills: digital inequality among older adults. Univ Access Inf Soc 2018;18(4):881-890. [doi: 10.1007/s10209-018-0617-5]
89. Sun X, Yan W, Zhou H, Wang Z, Zhang X, Huang S, et al. Internet use and need for digital health technology among the elderly: a cross-sectional survey in China. BMC Public Health 2020;20(1):1386 [FREE Full text] [doi: 10.1186/s12889-020-09448-0] [Medline: 32917171]
90. Hu B, Liu X, Lu C, Ju X. Prevalence and intervention strategies of health misinformation among older adults: a meta-analysis. J Health Psychol 2025;30(7):1427-1443. [doi: 10.1177/13591053241298362] [Medline: 39607815]
91. Vivion M, Reid V, Dubé E, Coutant A, Benoit A, Tourigny A. How older adults manage misinformation and information overload - a qualitative study. BMC Public Health 2024;24(1):871 [FREE Full text] [doi: 10.1186/s12889-024-18335-x] [Medline: 38515081]
92. Bawden D, Robinson L. Information overload: an overview. In: Oxford Encyclopedia of Political Decision Making. England: Oxford University Press; 2020.
93. Shahrzadi L, Mansouri A, Alavi M, Shabani A. Causes, consequences, and strategies to deal with information overload: a scoping review. Int J Inf Manag Data Insight 2024;4(2):100261. [doi: 10.1016/j.jjimei.2024.100261]
94. Zou Y, Sun K, Afnan T, Abu-Salma R, Brewer R, Schaub F. Cross-contextual examination of older adults' privacy concerns, behaviors, and vulnerabilities. PoPETs 2024;2024(1):133-150. [doi: 10.56553/popets-2024-0009]
95. Mihelič KP, Bernik I, Vrbovec S, Mihelič A. Perceived threat of cyber attacks and its role in the adoption of tablet computers by older adults. 2023 Presented at: Proceedings of the Central European Conference on Information and Intelligent System; 2023 May 29; Dubrovnik, Croatia p. 49-56. [doi: 10.1145/3590777.3590814]
96. Savitz DA, Wellenius GA. Can cross-sectional studies contribute to causal inference? It depends. Am J Epidemiol 2023;192(4):514-516. [doi: 10.1093/aje/kwac037] [Medline: 35231933]
97. Kelly SE, Brooks SP, Benkhedda K, MacFarlane AJ, Greene-Finestone LS, Skidmore B, et al. A scoping review shows that no single existing risk of bias assessment tool considers all sources of bias for cross-sectional studies. J Clin Epidemiol 2024;172:111408 [FREE Full text] [doi: 10.1016/j.jclinepi.2024.111408] [Medline: 38844117]
98. Mubarak F, Nycyk M. Teaching older people internet skills to minimize grey digital divides. JICES 2017;15(2):165-178. [doi: 10.1108/jices-06-2016-0022]
99. Zhang M. Older people's attitudes towards emerging technologies: a systematic literature review. Public Underst Sci 2023;32(8):948-968 [FREE Full text] [doi: 10.1177/09636625231171677] [Medline: 37204075]
100. Lu X, Yao Y, Jin Y. Digital exclusion and functional dependence in older people: findings from five longitudinal cohort studies. EClinicalMedicine 2022;54:101708 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101708] [Medline: 36353265]
101. Paez A. Gray literature: an important resource in systematic reviews. J Evid Based Med 2017;10(3):233-240. [doi: 10.1111/jebm.12266] [Medline: 28857505]
102. Sieck CJ, Sheon A, Ancker JS, Castek J, Callahan B, Siefer A. Digital inclusion as a social determinant of health. NPJ Digit Med 2021;4(1):52 [FREE Full text] [doi: 10.1038/s41746-021-00413-8] [Medline: 33731887]


## Abbreviations

**CGSS:** Chinese General Social Survey
**CHARLS:** China Health and Retirement Longitudinal Study
**ICT:** information and communication technology
**JBI:** Joanna Briggs Institute
**PRISMA-S:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Searching
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
**SRH:** self-rated health



XSL•FO
RenderX

*Edited by M Balcarras; submitted 04.May.2025; peer-reviewed by G Amabili, F Fukuzawa; comments to author 11.Nov.2025; revised version received 07.Jan.2026; accepted 21.Jan.2026; published 19.Feb.2026.*

*Please cite as:*
*AboJabel H, Abo-Rass F*
*Internet Use and Self-Rated Health Among Older Adults: Scoping Review*
*Interact J Med Res 2026;15:e76930*
*URL: https://www.i-jmr.org/2026/1/e76930*
*doi:10.2196/76930*
*PMID:*

©Hanan AboJabel, Fareeda Abo-Rass. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 19.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Review

# Treating Behavioral Addictions With Augmented Reality and Virtual Reality: Scoping Review

Felicia Xin Rou Chiew[1*], BSc(OT); Genevieve Huimin Li[2*], BSc(OT); Victrine Tseung[3], MSc (OT), PhD; Jing Shi[4*], BHSc (Hons), MSc (OT), PhD

[1]Occupational Therapy, National University Health System, Singapore, Singapore

[2]Occupational Therapy, Tan Tock Seng Hospital, Singapore, Singapore

[3]Brain and Heart Nexus Research Program, University of Ottawa Heart Institute, Ottawa, ON, Canada

[4]Health and Social Sciences, Singapore Institute of Technology, Singapore, Singapore

[*]these authors contributed equally

**Corresponding Author:**
Jing Shi, BHSc (Hons), MSc (OT), PhD
Health and Social Sciences
Singapore Institute of Technology
1 Punggol Coast Road
Singapore, 828608
Singapore
Phone: 65 65924288
Email: j.shi@alumni.utoronto.ca

## Abstract

**Background:** The use of augmented reality (AR) and virtual reality (VR) to address addictive behaviors such as substance use disorders and gambling disorders has been growing. However, little has been done to explore the use of AR and VR in the treatment of other behavioral addictions.

**Objective:** This scoping review aims to provide an overview of existing literature on AR and VR interventions for behavioral addictions. Specifically, the research questions are as follows: (1) What behavioral addictions or behavioral harms are being treated using AR and/or VR? (2) What AR and/or VR treatment interventions are being used to treat these behavioral addictions?

**Methods:** This scoping review was conducted based on the framework first proposed by Arksey and O'Malley, later refined by Levac et al, and further outlined in the Joanna Briggs Institute (JBI) Manual for Evidence. The literature was searched in the following databases: CINAHL, PsycArticles, PsycInfo, PubMed, and Web of Science, with Google advanced search complementing the search on Feb 22, 2023. Studies were screened by 2 independent reviewers based on inclusion criteria (all ages; any behavioral addiction, problematic behavior, or behavioral harm; AR or VR treatments and interventions) and exclusion criteria (pornography, sexual, and paraphilic disorders). Discrepancies were resolved by third and fourth reviewers. As this study is a scoping review, risk of bias was not assessed. Data were extracted and presented in tabular form as well as through conceptual analysis as a narrative summary.

**Results:** A total of 9 studies were included in this review, 4 studies on video gaming and 5 studies on gambling behaviors. Participants' age ranged from 12 to 65 years. Only the use of VR was identified. VR was used as a platform for cue exposure therapy and skills training in both gaming and gambling disorders. VR therapy was effective alone or in combination with other treatments and was comparable to traditional interventions. No adverse effect was reported in the studies.

**Conclusions:** VR is efficacious in treating behavioral addictions and can replace or be used in conjunction with traditional methods. Future directions include using VR with other psychotherapy or relapse prevention, applying VR to treat other addictions, and investigating harmful side effects of VR use. The frequency and duration of sessions can also be optimized. A limitation of this study is that there may be other documents beyond those published and searched in gray literature that could not be included in this review due to time and resource restrictions. The use of AR in the treatment of behavioral addictions did not yield any results in this review. However, VR application in behavioral addiction is promising, potentially efficacious, and capable of multiple applications.

**International Registered Report Identifier (IRRID):** RR2-10.4309/EGCN4808



XSL•FO
RenderX

*(Interact J Med Res 2026;15:e77011)* doi:10.2196/77011

**KEYWORDS**

behavioral addictions; virtual reality; augmented reality; intervention; video gaming; gambling; scoping review; behavioral harms; treatment; cue exposure

# Introduction

## Background

Behavioral addictions are characterized by repetitive and rewarding behaviors that disrupt an individual's life [1]. Only 2 behavioral "addictions"—gambling disorder and internet gaming disorder—are acknowledged in the *ICD-11* (*International Classification of Diseases, 11th Revision*) [1]. Substance use disorders and compulsive sexual behavior disorders are excluded, as they are classified separately under the *ICD-11* due to their distinct clinical profiles. Other behavioral harms such as exercise-related harms [2], mobile phone harms [3], and social media harms [4] are also widely debated. Systematic reviews and meta-analyses indicate that the worldwide prevalence of gaming disorders is approximately 3.05% [5], while the prevalence of gambling disorder is approximately 1.29% [6]. These findings show similarities to the prevalence of certain substance use disorders, estimated at 2.6% [7]. The World Health Organization (WHO) presentations on addictions at international forums in 2017 and 2019 [8] highlight an increase in attention on behavioral addictions.

Both recognized behavioral addictions have been linked to negative consequences. Internet gaming disorders have been associated with issues such as lower academic achievement, sleep loss [9], difficulties with interpersonal relationships, and effects on career opportunities [10]. While there exist studies that highlight potential benefits of gaming, especially for open-world games [11,12], this study focuses on excessive or problematic gaming behaviors that extend beyond healthy recreational use. These patterns, when persistent, may lead to significant psychosocial and physical difficulties. Gambling disorders may negatively impact personal relationships [13], employment, fitness, sleep [14], and overall health and well-being [15]. Recently, there has been a significant rise in the development of preventative strategies, assessment tools, and treatment methods for gaming and gambling disorders. Other potentially problematic behaviors that are less studied, such as social media use, shopping [16], exercise [17], work [18], general internet use [19], and other behaviors that do not involve substance intake, may also negatively impact psychosocial and physical well-being.

Current treatments like cognitive behavioral therapy (CBT) have proven effective in lessening problematic behaviors by empowering clients with motivational and cognitive tools to initiate change and avoid relapse [20]. Additionally, CBT helps individuals develop skills for regulating their emotions and problem-solving [20]. Other common interventions include mindfulness-based cognitive therapy to help with managing cravings [21] and family therapy and support groups to provide social support for patients [22].

## Technological Advances in Treatment Modalities

Mental health interventions have recently welcomed new innovations and experiences that showcase new technologies [23]. Two of these technologies are augmented reality (AR) and virtual reality (VR). AR technology establishes a computer-generated setting that produces a unique reality where real and virtual elements are perceived simultaneously and interactively [24]. One therapeutic application of AR technology is an AR-powered smart training program designed to teach tooth brushing and enhance oral hygiene for individuals with intellectual disabilities [25]. As a more advanced technology, VR uses computer graphics, motion sensors, and 3D display technologies that fully immerse users [26]. VR technology has shown positive outcomes in enhancing functional performance among adults in psychosocial rehabilitation [27]. Until recent years, VR faced limitations due to its high cost and poor quality of graphics and content available. However, growing interest in VR from the video game industry has made VR more accessible by reducing costs and increasing capabilities, resulting in an increasing uptake of VR technology [28].

The use of AR and VR to address mental health issues such as anxiety, substance use disorders, and gambling disorders has also been growing. AR and VR have been used in exposure therapy, where positive results were shown. One study found that an AR exposure app with graded tasks, varying in intensity and duration, effectively reduced the participants' subjective fear and disgust toward spiders [29]. Another study found that VR therapy effectively decreased craving and improved self-control in individuals with substance use disorders by exposing them to stimulating cues, ranging from pictures of items to immersive environments [30]. In addition, VR has proven helpful in the treatment of anxiety through its ability to generate virtual environments that induce anxiety, serving as a form of exposure therapy. Conducting exposure therapy in a virtual setting instead of in real life lowers the risk of worsening symptoms that might occur with direct exposure, making it a safer approach [31]. VR has also been used in providing role-playing opportunities for vocational training for people diagnosed with schizophrenia. Participants who received a 10-hour virtual job interview training had higher blind-rated scores by human resource members and higher self-rated confidence levels [32]. While there is growing interest in AR and VR as treatment modalities, research exploring the use of AR and VR in the treatment of behavioral addictions has not been consolidated. Hence, this scoping review aims to provide a comprehensive overview of existing literature on the use of AR and VR treatment for behavioral addictions or behavioral harms and to synthesize knowledge gaps for future research.



XSL•FO
RenderX

## Methods

### Overview

A scoping review was performed systematically according to the methodological framework proposed and augmented by Arksey and O'Malley [33], Levac et al [34], and the Joanna Briggs Institute (JBI) Manual for Evidence Synthesis [35]. The scoping review framework has 6 stages as described in the subsequent sections. The protocol for this study was first submitted for publication on January 27, 2023, and published after peer review [36].

### Stage 1: Identifying Research Questions

This scoping review aims to provide a comprehensive overview of existing literature on the use of AR and VR treatment for behavioral addictions or behavioral harms and synthesize knowledge gaps for future research. Our review questions were as follows:

1. What behavioral addictions or behavioral harms are being treated using AR and/or VR?
2. What AR and/or VR treatment interventions are being used to treat behavioral addictions?

### Stage 2: Identifying Relevant Studies

A comprehensive search was conducted in 5 databases, including CINAHL, PsycArticles, PsycInfo, PubMed, and Web of Science. Additionally, gray literature and Google advanced search were searched (the first 10 pages were screened). The search strategy aimed to identify studies published in English in the last 10 years where full-text publications were available. This was to account for the validity of recent technological advancements and to include a larger number of relevant articles that can be used in this scoping review.

To identify relevant literature, the selected databases above were searched using keywords after consultation with an academic librarian (see protocol [36] for a full list of keywords used). Boolean operators were applied to combine search terms logically. "OR" was used to group synonyms, ensuring the inclusion of all relevant terms. "AND" was used to combine the main keywords, such as "behavioral addiction," "virtual reality," and "intervention," so that only studies addressing all 3 areas were retrieved. Quotation marks were placed around multiword phrases to search for exact matches, while truncation symbols were used to capture variations in word endings. New keywords found in identified articles were included in the final searches across all databases. The full list of search terms used can be found in Multimedia Appendix 1. Finally, reference lists of all articles and reports were hand-searched for additional studies to ensure comprehensive identification of literature [35].

### Stage 3: Study Selection

#### *Inclusion and Exclusion Criteria*

Studies were selected based on the inclusion and exclusion criteria to meet the scope of the inquiry. The literature included in this review consists of published articles in peer-reviewed journals using all types of methodologies, including protocols. Gray literature is also included to minimize publication bias through Google advanced search. Only literature published in English was screened, reflecting the language expertise of the reviewers; nevertheless, sources of evidence from any country were eligible for inclusion. Inclusion criteria for participants were studies that involved people who exhibit behavioral addictions at all ages. Concepts of AR and VR interventions that can be used in treating behavioral addictions were included, such as assessments, perspectives, and outcomes measured for these interventions. Exclusion criteria were reviews and conference abstracts. Research that studied pornography addiction and compulsive sexual behaviors was excluded due to separate categorizations of compulsive sexual behavior disorder and paraphilic disorders from the *ICD-11* [1].

#### *Study Identification*

The study selection consisted first of screening titles and abstracts according to the inclusion and exclusion criteria by 2 independent reviewers (FXRC and GHL). If titles and abstracts matched the inclusion and exclusion criteria of the scoping review, a full-text review commenced. Two reviewers (FXRC and GHL) worked independently to review and extract relevant articles before importing them into Zotero (Corporation for Digital Scholarship), an open-source reference management software that eliminated duplicates of citations retrieved from multiple databases before screening [37]. A third and fourth reviewer (JS and VT) were consulted for ambiguities until a consensus was reached. Reasons for the exclusion of full-text studies that did not meet the inclusion criteria were recorded and reported in this scoping review. The search and screening processes were presented in a PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flow diagram [38].

### Stage 4: Charting the Data

Data were extracted from selected articles and recorded in a data extraction table in accordance with the methodology of scoping reviews [35]. The data extraction table was used to collate the extracted data based on identified data fields. These data fields include type of behavioral addiction, aim of study, treatment details, control groups, outcome measures, results, future directions, and conclusion. The first 2 authors (FXRC and GHL) independently extracted data from all the full texts. The last 2 authors (JS and VT) then cross-checked all the extractions.

### Stage 5: Collating, Summarizing, and Reporting the Results

Results are described and collated according to the PRISMA-ScR checklist [38] and the JBI methodological guide for scoping reviews [35] (Multimedia Appendix 2). A conceptual analysis was then carried out to identify and synthesize the key findings emerging from the included studies into main themes. Common characteristics of included studies were identified, and then comparisons were made between these studies under each identified concept. These were presented as a narrative summary.

### Stage 6: Consultation Exercise

The sixth stage is an optional consultation stage of scoping reviews [33]. The purpose of this stage is to gather additional



XSL•FO
RenderX

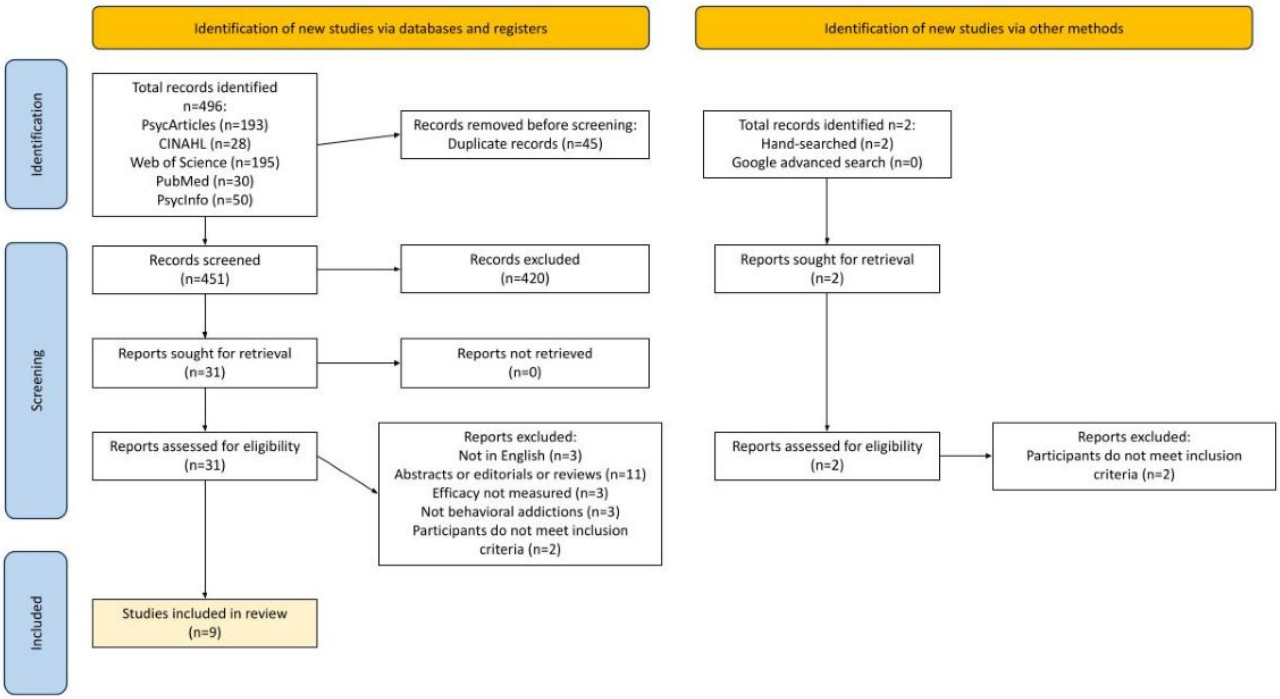
insights, perspectives, or references from stakeholders. Furthermore, this stage assesses whether findings resonate with real-world experiences to enhance the relevance and applicability of this scoping review. For this review, we reached out to 6 stakeholders—identified based on the research team's established professional networks—including academics, digital developers, and clinicians, about our findings on the identified preliminary literature findings from stage 5. A set of questions, along with a draft of this scoping review manuscript, was emailed to identified stakeholders and consultants for their feedback.

# Results

## Search Results

Five databases were searched on February 2, 2023. An overview of the data identification and selection process can be found in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart (Figure 1). A total of 496 studies were retrieved, of which 45 were duplicates and 420 did not meet the inclusion criteria. In total, 31 full-text studies were assessed for eligibility, which resulted in 9 studies [39-47] for inclusion in this scoping review.

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.




## Study Characteristics

Four studies [43-46] were on video gaming, while 5 studies [39-42,47] explored gambling behaviors. All 9 studies explored the use of VR as an intervention, while no studies used AR. The reviewed articles were from Canada (n=3) [39,41,47], Turkey (n=1) [46], Italy (n=1) [40], and South Korea (n=4) [42-45]. Among the studies that provided information on the age range (n=5) [39,40,44,46,47], they covered ages from 12 to 65 years. Studies on gambling behaviors [39-42,47] recruited participants aged 18-65 years, while studies on gaming behaviors [43-46] recruited younger participants aged 12-25 years. Furthermore, when it came to the sex of the participants recruited, 5 of the studies [42-46] exclusively recruited male participants. One study [47] recruited participants who were mainly White, while another [45] carried out its study on Asian participants. The other 7 studies [39-44,46] did not state the ethnicity of participants. Overall, participants across these studies represented a wide age range and were predominantly male. Among the 9 studies, 6 [39,41-43,45,47] were experimental designs, 2 [44,46] were randomized controlled trials, while one [40] was a study protocol for an experimental design. The characteristics and research aims of the 9 articles are presented in the extraction table titled, "Summary of virtual reality applications in behavioral addictions" (Multimedia Appendix 3).

Four themes were identified in the included studies: (1) the application of VR is predominantly for reducing urges, (2) VR therapies are generally efficacious compared to other treatments, (3) adverse effects are not found in the varied uses of VR treatment for behavioral addictions, and (4) further research is needed to support the use of VR in behavioral addictions treatment.

## Applications of VR Are Predominantly for Reducing Urges

Across all 9 studies, VR was used either in the treatment of gambling (n=5) or gaming disorders (n=4). No other types of behavioral harms were found. The application of VR varied across the studies. VR was mostly used as a platform for cue exposure therapy (CET) [39-44]. CET is a behavioristic psychological approach where individuals are deliberately exposed to relevant addictive cues with the aim of extinguishing conditioned responses associated with the cue by systematically pairing them with a lack of the addictive substance or behavior [48,49]. When CET is conducted through VR, users are



XSL•FO
RenderX

immersed in an environment where they are exposed to activity-related environmental stimuli. Giroux et al [41] used a virtual environment featuring 5 video lottery terminals along with other gambling-associated stimuli, including people who are gambling on video lottery terminals, a billiard table, and alcoholic beverages. Similarly, Park et al [42] included gambling-related cues such as navigating and playing a casino game, discussing gambling, and witnessing a jackpot scene. Park et al [43] exposed participants to gaming cues that they frequently play in video games, such as scenes of battling, leveling up, or shooting. Whereas Shin et al [44] exposed participants to scenes of video game invitations, the internet café entrance, and conversations on gaming.

VR was also used as a medium to teach specific skills, such as refusal skills and skills on managing game-related family conflicts. This was done by adopting VR as an environment for participants to engage in a role-playing situation to practice the aforementioned skills [44,45]. Shin et al [44] had participants reject gaming invitations through practicing refusal skills. These participants were told to refuse the avatars' invitations to play together in a game. In contrast, Shin et al [45] targeted resolving and minimizing game-related conflict with family through risk-benefit analysis and perspective-taking exercises. In the risk-benefit analysis, participants were asked to consider the benefits and disadvantages of halting the game immediately and to evaluate the impact of each example on them using a visual analog scale embedded in the VR program. In the perspective-taking exercise, they were asked to voice out the parents' behavior from their own perspective, then think about it from the parents' perspective. They were subsequently tasked to rate the extent to which they were able to understand the parents' perspective using the same visual analog scale. VR offers a realistic setting by which to practice real-life scenarios using different avatars and visual environments.

One study adopted VR to play sports games to study the therapeutic effects of VR therapy and aerobic training on fitness, physical activity, and anxiety in those with gaming disorder [46]. The VR therapy group played different sports games, including boxing, hurdles, and beach volleyball, on an Xbox Kinect 360 (Microsoft Corp). The games involved simultaneous body movements of the participants, including jumping and hitting movements. This helped to encourage more physical activity and prevent the worsening of gaming disorder [46]. The last study used VR to evaluate the effect of gambling goals on chasing behavior [47]. Chasing behavior refers to the persistent pursuit of losses through continued gambling, involving a pattern of escalating wagers and increased risk-taking in an effort to recoup money lost during previous gambling activities [47]. Participants played on slot machines and had to decide whether to stop or continue after several rounds. Their behaviors were tracked and compared with their gambling goals, which were assessed through a questionnaire (eg, how much they wanted to win and how motivated they were to win).

## VR Therapies Are Generally Efficacious Compared to Other Treatments

The studies in this scoping review either examined the efficacy of VR alone or in comparison to traditional CBT. Comparing pre- and post-effects of VR therapy, VR therapy was found to be effective in reducing the severity of behavioral addictions [39,42-46]. For gaming disorders, the VR therapy group had significantly shorter gaming time ($P$=.04) and increased weekly physical activity values ($P$=.03) as compared to a control group that did not receive any treatment [47] and showed significantly reduced scores on the Young's Internet Addiction Scale (YIAS; $P$<.01) after the treatment [43]. Practicing refusal skills during VR sessions also significantly lowered craving responses ($P$<.03) [44]. The Internet Gaming Disorder group had a higher frequency of halting the game after the risk/benefit assessment ($P$=.003) and anger expression task ($P$=.001) practices using VR as compared to baseline [45].

Similarly, for gambling disorders, VR therapy was effective in controlling gambling problems ($P$<.05) [39], and participants in the VR therapy group showed significant decreases across sessions in the subjective urge to gamble as compared to baseline ($P$=.04) [42]. However, one study found no reduction in urges to gamble after a single session of CET in a virtual gambling environment session [41]. It was speculated by the author that this could be due to the study having only one session of 20 minutes as compared to other studies with multiple sessions and longer durations. Overall, all but one study found promising results for the use of VR in the treatment of behavioral addictions.

When compared to other treatment modalities, VR therapy was found to be an effective and viable alternative [39,43,46]. Both CBT and VR therapy resulted in a reduction in the urge to gamble, with similar results regardless of the treatment modality ($P$<.001) [37]. The VR therapy program was shown to be as effective as the aerobic exercise program in reducing the severity of gaming disorder and anxiety, with both groups having significantly shorter gaming time ($P$=.04 and $P$=.04) and sedentary time ($P$<.001 and $P$=.03) [46]. Similarly, both CBT and VR therapy groups showed significant effects in reducing YIAS scores with no significant difference in YIAS score change between the CBT and VR therapy groups ($P$=.52) [43]. VR therapy was also effective when used in combination with CBT ($P$<.05) or relaxation exercises ($P$=.04) [39,42]. Furthermore, VR therapy demonstrated effectiveness in reducing the severity of online gaming addiction among individuals with attention deficit/hyperactive disorder symptoms by enhancing their engagement through greater immersion in the therapeutic process using VR. This reduced the challenge of maintaining their attention with conventional therapeutic techniques [43]. Two studies did not evaluate the efficacy of their treatments [40,47].

## Varied VR Treatment Programs With No Adverse Effects Reported

Across the 8 studies [39,41-47], with one [40] protocol study excluded, the treatment program varied in terms of type, frequency, duration, and the personnel who conducted the treatment (Table 1). In addition to investigating the use of VR therapy for the treatment of behavioral addictions, some studies [39,41,44,45] considered possible adverse effects, such as cybersickness, also known as simulator sickness. Cybersickness is a syndrome akin to motion sickness and is a potential side



XSL•FO
RenderX

effect during and/or after exposure to various virtual reality environments. While motion sickness is triggered by actual movement, cybersickness is triggered by visual stimulation [50].

**Table 1.** Program details and cybersickness measurement.

| Author (year of publication) | Exposure duration | Facilitated by | Measured cybersickness |
|---|---|---|---|
| Bouchard et al (2017) [39] | 20 min | Therapists | Yes |
| Giroux et al (2013) [41] | 20 min | Researcher | Yes |
| Lister et al (2016) [47] | 5 min | Experimenter | No |
| Maden et al (2022) [46] | 30 min | Physiotherapist | No |
| Park et al (2015) [42] | 40 min, 5 separate scenes | Not stated | No |
| Park et al (2016) [43] | 25 min | Psychiatrist | No |
| Shin et al (2018) [44] | Not stated | Not stated | Yes |
| Shin et al (2021) [45] | 2 scenes, 10-15 minutes each | Therapist | Yes |

In terms of duration, 5 studies had participants exposed to VR for at least 20 minutes [39, 41, 42, 43,46], while one study had participants in the VR environment for 5 minutes [47]. In one study, participants were exposed to 5 different scenes in 40 minutes [42]; however, the authors did not state if there were any breaks between the scenes. Similarly, a different study had participants in 2 different scenarios of about 10-15 minutes each [45], without indicating if a break was given in between. One study did not state the duration of virtual environment exposure [44]. Of the studies that specified treatment details, the number of treatments varied from one session [41] to 18 sessions [46].

Despite the varied frequency and duration of treatment, VR used in the treatment for behavioral addictions was not found to induce significant cybersickness. To measure the presence of cybersickness due to VR therapy, 3 studies used the cybersickness questionnaire [39,44,45], while one study used a post-VR exposure questionnaire [41]. The recorded levels of cybersickness did not increase in intensity compared to before VR therapy [39] or remained lower than established norms [44,45], affirming the feasibility of VR immersion for treating internet gaming disorder [44,45] and gambling disorder [39,41]. The other 5 studies [40,42,43,46,47] did not specify measuring cybersickness and did not report any negative side effects associated with VR therapy.

## Further Research Is Needed to Support the Use of VR in Behavioral Addiction Treatment

The findings from the included studies in this scoping review highlighted the potential for further exploration of VR in the treatment of behavioral addictions and suggested expanding the use of VR within and beyond the scope of behavioral addictions. All studies should report facilitators of the programs to ensure that adequately trained personnel can administer the intervention. Of the studies that reported this, VR therapy was facilitated by therapists (n=2), psychiatrists (n=1), physiotherapists (n=1), and the researchers themselves (n=2). Two articles did not state the facilitators. Additionally, 2 studies called for improving the efficacy of VR for treatment to address gambling addictions. Park et al [42] suggested investigating the optimal frequency and duration of VR sessions. Similarly, for gaming addictions, Shin et al [44] stated the need to focus on developing VR treatments for internet gaming disorders and comparing them with conventional treatment protocols.

Four studies suggested additional research to assess the effectiveness of combining VR with other types of therapy. Bouchard et al [39] suggested using VR with other forms of psychotherapy or mindfulness. Giroux et al [41] recommended combining VR treatment and relapse prevention with cognitive intervention for gambling problems. Shin et al [45] expressed interest in examining the long-term effects of skill learning in addressing gaming disorders when VR is combined with CBT. Giordano et al [40], whose study protocol aimed to compare combined VR therapy and CBT with traditional CBT, indicated a need for future studies to include a nontreatment control group to better analyze the efficacy of the combined intervention. Beyond behavioral addictions, Bouchard et al [39] recommended exploring the potential of VR applications in addressing other types of addiction. They also cautioned that there is a need for research on the potential of VR addiction [39].

Outside of using VR, Lister et al [47] suggested discovering if there is a better way to reduce problematic gambling outcomes (chasing and losing money), such as redefining goal-setting for gambling achievements. They suggested that setting realistic gambling goals might be more effective in reducing problem gambling compared to the current limit-setting tools such as limiting time or money. Park et al [43] and Maden et al [46] did not mention future research directions.

## Consultation Response

Consultation questions, along with an earlier draft of our manuscript, were sent to 6 stakeholders; however, only one responded after the initial email and 2 follow-up emails. Due to the low response rate, no formal analysis was conducted for this stage. The questions and responses are available in Multimedia Appendix 4. Our consultant (Dr. Bhing Leet Tan) has experience working as an occupational therapist and more than 11 years of experience as the head of the department at a mental health hospital. She has also published research on using VR for rehabilitation and AR for people with intellectual and developmental disabilities. She noted that while younger clients may find VR interventions more acceptable due to familiarity, older individuals, particularly those with gambling addictions,



XSL•FO
RenderX

may require adjustment. She suggested grounding interventions within established frameworks like CBT or other frames of reference and providing interventions in multiple languages for psychoeducation. While she found it noteworthy that VR interventions showed positive outcomes in conflict resolution and assertiveness training, this did not come as a surprise to her due to VR's prior effectiveness in social skills training. Nonetheless, she identified gaps in the literature, including specific VR hardware used and the sustainability of urge reduction gains over time.

## Discussion

### Principal Findings

This scoping review was conducted to provide an overview of existing literature on the use of AR and VR treatment for behavioral addictions. VR was found to be used for the treatment of problem video gaming and problem gambling. However, no AR studies were found. A breadth of studies from 4 countries and participant age ranges (children to older adults) were found; however, participants in the gaming studies were younger than participants in gambling studies. All 9 studies explored either problematic gaming or gambling, with study participants being mostly male, and ethnicity was rarely reported.

CET and skills training were the most common uses of VR in treatment to help with controlling urges and for role-playing to practice specific skill sets. Additionally, it was used as a platform to encourage a healthier lifestyle by introducing alternative activities. VR was applicable for both gaming and gambling disorders. Overall, VR therapy shows promise in the treatment of both gaming and gambling disorders, either as a standalone intervention or in combination with other treatments. It has also shown comparable results to other treatment modalities, showcasing its potential as a viable alternative to traditional treatment approaches. Nevertheless, these findings should be interpreted with caution, as the current body of evidence is still preliminary. A comprehensive systematic review assessing the methodological quality of existing studies is warranted.

Furthermore, most studies in this review tested for adverse effects, but none were identified. Overall, the results indicated that regardless of the treatment type, participants were able to spend up to 40 minutes in the VR environment without experiencing any adverse effects. Additionally, the varied facilitators did not seem to have an impact on cybersickness experienced by participants.

### Comparison to Prior Work

Compared to previous literature, this scoping review both supports and extends prior findings in several ways. Consistent with earlier reviews, this review confirms that most research on VR interventions for behavioral addictions is concentrated in a small number of countries [51], notably, Canada (33.3%) and South Korea (44.4%) for this study [28]. Furthermore, 75% of the articles on gaming disorder are from South Korea, possibly due to esports being a prominent culture in South Korea [52]. South Korea also has a very high proportion of gamers among the adolescent population (as high as 91.5%) [53]. This review identified an acute gap in research on the use of AR as an intervention for behavioral addictions, notwithstanding its emerging promise in other fields in mental health rehabilitation [54-56] and in substance use disorders [57,58].

While the practicality of VR in clinical use could be limited by financial constraints [28] and higher production costs [59], AR is an economical and accessible alternative with easy incorporation into smart phones and mobile devices without the need for additional equipment such as headsets for VR environments [60]. Additionally, the ecological validity is higher in AR, as the problematic behavior is embedded in the real environment where the user is allowed to use their own body to interact with virtual objects as opposed to participating via a virtual representation of their body in VR [23]. A possible explanation for the lack of AR treatment approaches could be due to a limited understanding of AR by health care professionals and a lack of awareness of its potential benefits. There may also be more ethical concerns when using AR since it incorporates real-world environments into treatment, potentially making cue induction too realistic with no gold-standard precautions in place to manage clients.

CET and skills training were the most common uses of VR in treatment. These showed mainly positive results, not only in being effective in reducing the symptoms and severity of the addictions but also being comparable to traditional treatment modalities such as CBT. This aligns with the results from studies of patients with panic disorder and agoraphobia where VR exposure therapy was found to be effective and equivalent to CBT alone [61]. In fact, a randomized controlled trial found that self-led VR-based CBT was also found to be effective for panic disorder [62]. Although many other studies using VR, most commonly in the form of CET, found decreased craving and dependence in people with substance use disorders [28], an earlier study found no significant change in craving after VR exposure [63]. Overall, the potential of the use of VR in the treatment of behavioral addictions is promising, especially in combination with CBT. Further research, however, is needed to establish best practices.

A majority of the studies in this review that tested for adverse effects used the Simulator Sickness Questionnaire to test for cybersickness, which is the most widely used measure of sickness in virtual environments [64]. It has been recognized that VR can cause discomfort and cybersickness after prolonged use [65]. However, no adverse side effects were reported in any of the identified studies. In 8 studies, the duration participants spent in the VR environment was within 20 minutes, aligning with studies that confirmed this duration of VR usage did not lead to adverse effects. One study indicated that the VR therapy was 40 minutes, spent in 5 separate scenes [42]. This longer duration also did not lead to adverse effects of using VR, specifically cybersickness. Therefore, in this scoping review, VR is deemed appropriate in the treatment of behavioral addictions, as no significant adverse effects have been reported.

Possible reasons for no reports of cybersickness may be that individuals with gaming and gambling disorders may be spending prolonged periods of time in front of screens, such that they have a higher tolerance to light movements in VR



XSL•FO
RenderX

devices. Dużmańska et al [66] suggested that cybersickness can be avoided when the VR users are allowed enough time to adapt to the simulator conditions. Hence, prolonged periods of screen time during gaming and gambling can be a form of adaptation to the screen rate and locomotion, thus reducing the severity of cybersickness. Nevertheless, studies should screen for participants with known motion sickness, light sensitivities, epilepsy, and schizophrenia.

### Future Directions

Overall, the studies included in this review showed that VR was a viable mode of delivering therapy for behavioral addictions (gaming and gambling) and agree that further research is required to optimize the use of VR in treatment to maximize its potential. There is a scarcity of research on the topic of AR and VR in the treatment of behavioral addictions. Future programs that use AR or VR should consider publishing details of their program (eg, what their program was for and what tasks participants were asked to complete) and their effectiveness. Overall, more studies focusing on using AR as a treatment modality for behavioral addictions could be explored.

There is potential for AR and VR to be used in other types of problematic behaviors other than video gaming and gambling. Investigating cue exposure may provide a useful starting point, given that it is the most extensively studied intervention within the existing literature. As Bouchard et al [39] mentioned, combining VR with other psychotherapies is another avenue to explore in future intervention development. An 8-week mindfulness-based program was found to be effective for individuals with gambling problems [67]. A self-directed mindfulness program could be integrated with AR technology in the form of an app to facilitate increased accessibility to treatment. Comparison studies of VR treatment with pharmacological treatments could also be conducted. The ability to configure virtual environments also allows for personalization of treatment, potentially enhancing the effectiveness and receptivity of therapeutic interventions. All in all, VR treatments appear to be efficacious. However, long-term studies and evaluation of the quality of evidence are needed to determine its viability.

This review's consultant points out that the potential of VR interventions among different age groups underscores its potential as a versatile therapeutic tool. She suggested the importance of integrating theoretical frameworks into VR interventions to enhance efficacy and address clients' needs effectively. Further research is warranted to address the identified gaps and refine VR intervention strategies for optimal clinical outcomes.

### Strengths and Limitations

As this study was a scoping review, we did not aim to evaluate the risk of bias and quality of the studies found. First, only 9 articles from a small range of countries were found, which greatly limits the amount of evidence in evaluating the actual potential and efficacy of VR in behavioral addictions and in the different contexts. As more articles are published, a future systematic review may be warranted to synthesize the quality of the literature on AR and VR treatments in behavioral addictions. Second, there may be other relevant documents beyond the published articles and gray literature searches, not available in the public domain or published in other languages on this topic, which were not found by our research team and which could not be included in this review due to time and resource restrictions. To overcome this, requests have been made to the library to access full-text articles, and references of included studies were hand-searched for relevant studies. However, the diversity of studies in this review was still limited. Systematic searches of hospital, government, and global agencies may have revealed more studies.

Third, additional limitations based on this review's findings are that only studies on the use of VR with gambling and gaming disorders were identified, despite the wide spectrum of problematic behaviors that are studied (eg, excessive social media use and online shopping). This limits conclusions that can be drawn about AR and VR for other types of behavioral harms relating to technology and internet use. Finally, there were different types of behavioral addictions and varying age groups and sexes of participants. Although it served the purpose of our review, the results may not be as straightforward in providing an accurate picture of the way AR and VR can be incorporated in effecting the treatment of behavioral addictions. On the other hand, this scoping review helps to consolidate available studies and provide an overview and potential directions that may enhance the use of AR and VR in behavioral addiction treatment.

### Conclusion

In this scoping review, we consolidated existing literature on the use of VR in the treatment of behavioral addictions, identifying types of behavioral addictions or harms that are being treated and the types of treatments or interventions that are conducted. This review shows that behavioral addictions in gaming and gambling can be effectively treated by using VR. For treatment, VR therapy can be used by itself or in conjunction with CBT for skills training and controlling urges. In research, VR can be used as a modality to simulate real events, such as to study gambling behaviors or the therapeutic effects of physical activities. The range of future research suggested by the studies shows that the full potential of VR therapy is yet to be realized. More research is needed to optimize VR treatment for behavioral addictions in terms of type, frequency, and duration. As further research progresses, this review hopes to encourage the continuous development of treatment options that advances with technology.

## Acknowledgments

We would like to acknowledge Prof Tan Bhing Leet for her thoughtful responses in the consultation phase of this scoping review. We would also like to thank Chng Zhi Quan for his diligent work in preparing this manuscript for publication. Finally, we would



XSL•FO
RenderX

also like to acknowledge support from Singapore Institute of Technology's library staff. The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content of this manuscript.

## Funding

Funding was received from an Ignition Grant at the Singapore Institute of Technology. The grantor did not have any role in the conceptualization or reported findings of this study.

## Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

## Authors' Contributions

Conceptualization: FXRC, GHL, JS
Data curation: FXRC, GHL, VT, JS
Formal analysis: FXRC, GHL, VT, JS
Funding acquisition: JS
Investigation: FXRC, GHL, VT
Methodology: FXRC, GHL, JS
Project administration: JS
Supervision: JS
Writing-original draft: FXRC, GHL
Writing-review & editing: JS, FXRC, GHL, VT

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms and keywords.
[DOCX File , 15 KB - ijmr_v15i1e77011_app1.docx ]

Multimedia Appendix 2
PRISMA-ScR checklist.
[DOCX File , 86 KB - ijmr_v15i1e77011_app2.docx ]

Multimedia Appendix 3
Summary of virtual reality applications in behavioural addictions.
[DOCX File , 3649 KB - ijmr_v15i1e77011_app3.docx ]

Multimedia Appendix 4
Consultation questions and responses.
[DOCX File , 27 KB - ijmr_v15i1e77011_app4.docx ]

## References

1. International statistical classification of diseases and related health problems. World Health Organization. 2019. URL: https://icd.who.int/ [accessed 2022-11-12]
2. Trott M, Johnstone J, Pardhan S, Barnett Y, Smith L. Changes in body dysmorphic disorder, eating disorder, and exercise addiction symptomology during the COVID-19 pandemic: a longitudinal study of 319 health club users. Psychiatry Res 2021;298:113831 [FREE Full text] [doi: 10.1016/j.psychres.2021.113831] [Medline: 33652248]
3. Bsisu I, Dweikat A, Abojaradeh A, Nael N, Hijazin N, Khalifeh AH, et al. Mobile phone addiction; an emerging problem with psychological impact on Arabian adults. Int J Soc Sci Res 2021;3(3):41-57.
4. Amirthalingam J, Khera A. Understanding social media addiction: a deep dive. Cureus 2024;16(10):e72499. [doi: 10.7759/cureus.72499] [Medline: 39600781]
5. Stevens MW, Dorstyn D, Delfabbro PH, King DL. Global prevalence of gaming disorder: a systematic review and meta-analysis. Aust N Z J Psychiatry 2021;55(6):553-568. [doi: 10.1177/0004867420962851] [Medline: 33028074]
6. Gabellini E, Lucchini F, Gattoni ME. Prevalence of problem gambling: a meta-analysis of recent empirical research (2016-2022). J Gambl Stud 2023;39(3):1027-1057. [doi: 10.1007/s10899-022-10180-0] [Medline: 36586057]



XSL•FO
RenderX

7. Degenhardt L, Glantz M, Evans-Lacko S, Sadikova E, Sampson N, Thornicroft G, World Health Organization's World Mental Health Surveys collaborators. Estimating treatment coverage for people with substance use disorders: an analysis of data from the world mental health surveys. World Psychiatry 2017;16(3):299-307 [FREE Full text] [doi: 10.1002/wps.20457] [Medline: 28941090]
8. Abbott M. Gambling and gambling-related harm: recent world health organization initiatives. Public Health 2020;184:56-59. [doi: 10.1016/j.puhe.2020.04.001] [Medline: 32402594]
9. Kim H, Ha J, Chang WD, Park W, Kim L, Im CH. Detection of craving for gaming in adolescents with internet gaming disorder using multimodal biosignals. Sensors (Basel) 2018;18(1):102 [FREE Full text] [doi: 10.3390/s18010102] [Medline: 29301261]
10. Schneider LA, King DL, Delfabbro PH. Family factors in adolescent problematic internet gaming: a systematic review. J Behav Addict 2017;6(3):321-333 [FREE Full text] [doi: 10.1556/2006.6.2017.035] [Medline: 28762279]
11. Anto A, Basu A, Selim R, Foscht T, Eisingerich AB. Open-world games' affordance of cognitive escapism, relaxation, and mental well-being among postgraduate students: mixed methods study. J Med Internet Res 2024;26:e63760 [FREE Full text] [doi: 10.2196/63760] [Medline: 39689301]
12. Arigayota A, Duffek B, Hou C, Eisingerich AB. Effects of the legend of Zelda: breath of the wild and Studio Ghibli films on young people's sense of exploration, calm, mastery and skill, purpose and meaning, and overall happiness in life: exploratory randomized controlled study. JMIR Serious Games 2025;13:e76522 [FREE Full text] [doi: 10.2196/76522] [Medline: 40750097]
13. Tulloch C, Browne M, Hing N, Rockloff M, Hilbrecht M. How gambling harms others: the influence of relationship-type and closeness on harm, health, and wellbeing. J Behav Addict 2023;12(3):697-710 [FREE Full text] [doi: 10.1556/2006.2023.00036] [Medline: 37450370]
14. Muggleton N, Parpart P, Newall P, Leake D, Gathergood J, Stewart N. The association between gambling and financial, social and health outcomes in big financial data. Nat Hum Behav 2021;5(3):319-326. [doi: 10.1038/s41562-020-01045-w] [Medline: 33542528]
15. Butler N, Quigg Z, Bates R, Sayle M, Ewart H. Gambling with your health: associations between gambling problem severity and health risk behaviours, health and wellbeing. J Gambl Stud 2020;36(2):527-538 [FREE Full text] [doi: 10.1007/s10899-019-09902-8] [Medline: 31705379]
16. Ibrahim NAN, Husseini F, Aisyah N, Camelia N, Khazlyn N, Faiz W, et al. Online shopping behaviour in youth: a systematic review of the factors influencing online shopping in young adults. Int J Acad Res Bus Soc Sci 2023;13(2):168-178. [doi: 10.6007/ijarbss/v13-i2/16257]
17. Colledge F, Sattler I, Schilling H, Gerber M, Pühse U, Walter M. Mental disorders in individuals at risk for exercise addiction - a systematic review. Addict Behav Rep 2020;12:100314 [FREE Full text] [doi: 10.1016/j.abrep.2020.100314] [Medline: 33364322]
18. Berta K, Pesthy ZV, Vékony T, Farkas BC, Németh D, Kun B. The neuropsychological profile of work addiction. Sci Rep 2023;13(1):20090 [FREE Full text] [doi: 10.1038/s41598-023-47515-9] [Medline: 37973989]
19. Shi J, van der Maas M, Yu L, Jiang Q, Agasee S, Turner NE. Current prevention strategies and future directions for problem internet use. Curr Opin Behav Sci 2022;48:101231. [doi: 10.1016/j.cobeha.2022.101231]
20. McHugh RK, Hearon BA, Otto MW. Cognitive behavioral therapy for substance use disorders. Psychiatr Clin North Am 2010;33(3):511-525 [FREE Full text] [doi: 10.1016/j.psc.2010.04.012] [Medline: 20599130]
21. Shonin E, Van Gordon W, Griffiths MD. Do mindfulness-based therapies have a role in the treatment of psychosis? Aust N Z J Psychiatry 2014;48(2):124-127. [doi: 10.1177/0004867413512688] [Medline: 24220133]
22. Bowden-Jones H, George S. A Clinician's Guide to Working with Problem Gamblers. London: Routledge; 2015.
23. Baus O, Bouchard S. Moving from virtual reality exposure-based therapy to augmented reality exposure-based therapy: a review. Front Hum Neurosci 2014;8:112 [FREE Full text] [doi: 10.3389/fnhum.2014.00112] [Medline: 24624073]
24. Bakır Ç, Abbas SO, Sever E, Özcan Morey A, Aslan Genç H, Mutluer T. Use of augmented reality in mental health-related conditions: a systematic review. Digit Health 2023;9:20552076231203649 [FREE Full text] [doi: 10.1177/20552076231203649] [Medline: 37791140]
25. Jeon B, Oh J, Son S. Effects of tooth brushing training, based on augmented reality using a smart toothbrush, on oral hygiene care among people with intellectual disability in Korea. Healthcare (Basel) 2021;9(3):348 [FREE Full text] [doi: 10.3390/healthcare9030348] [Medline: 33803836]
26. Velev D, Zlateva P. Virtual reality challenges in education and training. Int J Learn Teach 2017;3:33-37. [doi: 10.18178/ijlt.3.1.33-37]
27. Tan BL, Shi J, Yang S, Loh H, Ng D, Choo C, et al. The use of virtual reality and augmented reality in psychosocial rehabilitation for adults with neurodevelopmental disorders: a systematic review. Front Psychiatry 2022;13:1055204 [FREE Full text] [doi: 10.3389/fpsyt.2022.1055204] [Medline: 36590624]
28. Segawa T, Baudry T, Bourla A, Blanc JV, Peretti CS, Mouchabac S, et al. Virtual reality (VR) in assessment and treatment of addictive disorders: a systematic review. Front Neurosci 2019;13:1409 [FREE Full text] [doi: 10.3389/fnins.2019.01409] [Medline: 31998066]




XSL•FO
RenderX

29. Zimmer A, Wang N, Ibach MK, Fehlmann B, Schicktanz NS, Bentz D, et al. Effectiveness of a smartphone-based, augmented reality exposure app to reduce fear of spiders in real-life: a randomized controlled trial. J Anxiety Disord 2021;82:102442 [FREE Full text] [doi: 10.1016/j.janxdis.2021.102442] [Medline: 34246153]
30. Kim S, Kim E. The use of virtual reality in psychiatry: a review. Soa Chongsonyon Chongsin Uihak 2020;31(1):26-32 [FREE Full text] [doi: 10.5765/jkacap.190037] [Medline: 32612410]
31. Boeldt D, McMahon E, McFaul M, Greenleaf W. Using virtual reality exposure therapy to enhance treatment of anxiety disorders: identifying areas of clinical adoption and potential obstacles. Front Psychiatry 2019;10:773 [FREE Full text] [doi: 10.3389/fpsyt.2019.00773] [Medline: 31708821]
32. Smith MJ, Fleming MF, Wright MA, Roberts AG, Humm LB, Olsen D, et al. Virtual reality job interview training and 6-month employment outcomes for individuals with schizophrenia seeking employment. Schizophr Res 2015;166(1-3):86-91 [FREE Full text] [doi: 10.1016/j.schres.2015.05.022] [Medline: 26032567]
33. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
34. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
35. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. In: Aromataris E, Munn Z, editors. JBI Evid Synth. Australia: JBI Evid Synth; Oct 2020:2119-2126.
36. Chiew FXR, Li GH, Shi J. Virtual reality and augmented reality treatment for behavioural addictions: a scoping review protocol. J Gambl Issues 2025 Feb 21;56(suppl):61-73 [FREE Full text] [doi: 10.4309/EGCN4808]
37. Your personal research assistant. Zotero. 2024. URL: https://www.zotero.org/ [accessed 2026-02-20]
38. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
39. Bouchard S, Robillard G, Giroux I, Jacques C, Loranger C, St-Pierre M, et al. Using virtual reality in the treatment of gambling disorder: the development of a new tool for cognitive behavior therapy. Front Psychiatry 2017;8:27 [FREE Full text] [doi: 10.3389/fpsyt.2017.00027] [Medline: 28286486]
40. Giordano R, Donati MA, Zamboni L, Fusina F, Primi C, Lugoboni F. Alter game: a study protocol on a virtual "Serious Game" for relapse prevention in patients with gambling disorder. Front Psychiatry 2022;13:854088 [FREE Full text] [doi: 10.3389/fpsyt.2022.854088] [Medline: 35432033]
41. Giroux I, Faucher-Gravel A, St-Hilaire A, Boudreault C, Jacques C, Bouchard S. Gambling exposure in virtual reality and modification of urge to gamble. Cyberpsychol Behav Soc Netw 2013;16(3):224-231. [doi: 10.1089/cyber.2012.1573] [Medline: 23496679]
42. Park C, Park SM, Gwak AR, Sohn BK, Lee J, Jung HY, et al. The effect of repeated exposure to virtual gambling cues on the urge to gamble. Addict Behav 2015;41:61-64. [doi: 10.1016/j.addbeh.2014.09.027] [Medline: 25306387]
43. Park SY, Kim SM, Roh S, Soh M, Lee SH, Kim H, et al. The effects of a virtual reality treatment program for online gaming addiction. Comput Methods Programs Biomed 2016;129:99-108. [doi: 10.1016/j.cmpb.2016.01.015] [Medline: 26860055]
44. Shin YB, Kim JJ, Kim MK, Kyeong S, Jung YH, Eom H, et al. Development of an effective virtual environment in eliciting craving in adolescents and young adults with internet gaming disorder. PLoS One 2018;13(4):e0195677 [FREE Full text] [doi: 10.1371/journal.pone.0195677] [Medline: 29672530]
45. Shin Y, Kim J, Kim H, Kim S, Eom H, Jung YH, et al. Managing game-related conflict with parents of young adults with internet gaming disorder: development and feasibility study of a virtual reality app. JMIR Serious Games 2021;9(1):e22494 [FREE Full text] [doi: 10.2196/22494] [Medline: 33459603]
46. Maden C, Bayramlar K, Aricak OT, Yagli NV. Effects of virtual reality-based training and aerobic training on gaming disorder, physical activity, physical fitness, and anxiety: a randomized, controlled trial. Ment Health Phys Act 1992;23:100465. [doi: 10.1016/j.mhpa.2022.100465]
47. Lister JJ, Nower L, Wohl MJA. Gambling goals predict chasing behavior during slot machine play. Addict Behav 2016;62:129-134. [doi: 10.1016/j.addbeh.2016.06.018] [Medline: 27347654]
48. Hone-Blanchet A, Wensing T, Fecteau S. The use of virtual reality in craving assessment and cue-exposure therapy in substance use disorders. Front Hum Neurosci 2014;8:844 [FREE Full text] [doi: 10.3389/fnhum.2014.00844] [Medline: 25368571]
49. Zhang Y, Ndasauka Y, Hou J, Chen J, Yang LZ, Wang Y, et al. Cue-induced behavioral and neural changes among excessive internet gamers and possible application of cue exposure therapy to internet gaming disorder. Front Psychol 2016;7:675 [FREE Full text] [doi: 10.3389/fpsyg.2016.00675] [Medline: 27242589]
50. Davis S, Nesbitt K, Nalivaiko E. A systematic review of cybersickness. : ACM Press; 2014 Presented at: IE2014: Proceedings of the 2014 Conference on Interactive Entertainment; 2014 December 2 - 3; Newcastle, NSW p. E2014. [doi: 10.1145/2677758.2677780]




XSL•FO
RenderX

51. Tay JL, Xie H, Sim K. Effectiveness of augmented and virtual reality-based interventions in improving knowledge, attitudes, empathy and stigma regarding people with mental illnesses-a scoping review. J Pers Med 2023;13(1):112 [FREE Full text] [doi: 10.3390/jpm13010112] [Medline: 36675773]
52. Jin DY. Historiography of Korean esports: Perspectives on spectatorship. Int J Commun 2020;14:3727-3745. [doi: 10.5040/9781501368745.0009]
53. Kim D, Lee J. Addictive internet gaming usage among Korean adolescents before and after the outbreak of the COVID-19 pandemic: a comparison of the latent profiles in 2018 and 2020. Int J Environ Res Public Health 2021;18(14):7275 [FREE Full text] [doi: 10.3390/ijerph18147275] [Medline: 34299725]
54. Bridges SA, Robinson OP, Stewart EW, Kwon D, Mutua K. Augmented reality: teaching daily living skills to adults with intellectual disabilities. J Spec Educ Technol 2019;35(1):3-14. [doi: 10.1177/0162643419836411]
55. Chang YJ, Kang YS, Huang PC. An augmented reality (AR)-based vocational task prompting system for people with cognitive impairments. Res Dev Disabil 2013;34(10):3049-3056. [doi: 10.1016/j.ridd.2013.06.026] [Medline: 23880030]
56. de la Torre-Luque A, Valero-Aguayo L, de la Rubia-Cuestas EJ. Visuospatial orientation learning through virtual reality for people with severe disability. Int J Disabil Dev Educ 2017;64(4):420-435. [doi: 10.1080/1034912x.2016.1274022]
57. Brandon K, Vinci C, Kleinjan M, Hernandez L, Sawyer L, Sutton S, et al. Testing augmented reality for eliciting cue-provoked urges to smoke: toward moving cue-exposure into the real world. Nicotine Tob Res 2021;23(5):861-865 [FREE Full text] [doi: 10.1093/ntr/ntaa259] [Medline: 33277653]
58. Ghiţă A, Gutiérrez-Maldonado J. Applications of virtual reality in individuals with alcohol misuse: a systematic review. Addict Behav 2018;81:1-11. [doi: 10.1016/j.addbeh.2018.01.036] [Medline: 29421343]
59. Suso-Ribera C, Fernández-Álvarez J, García-Palacios A, Hoffman HG, Bretón-López J, Baños RM, et al. Virtual reality, augmented reality, and in vivo exposure therapy: a preliminary comparison of treatment efficacy in small animal phobia. Cyberpsychol Behav Soc Netw 2019;22(1):31-38 [FREE Full text] [doi: 10.1089/cyber.2017.0672] [Medline: 30335525]
60. Vinci C, Brandon KO, Kleinjan M, Brandon TH. The clinical potential of augmented reality. Clin Psychol (New York) 2020;27(3):e12357 [FREE Full text] [doi: 10.1111/cpsp.12357] [Medline: 33223628]
61. Pelissolo A, Zaoui M, Aguayo G, Yao S, Roche S, Ecochard R, et al. Virtual reality exposure therapy versus cognitive behavior therapy for panic disorder with agoraphobia: a randomized comparison study. J CyberTher Rehabil 2012;5(1):35-43 [FREE Full text]
62. Shin B, Oh J, Kim BH, Kim HE, Kim H, Kim S, et al. Effectiveness of self-guided virtual reality-based cognitive behavioral therapy for panic disorder: randomized controlled trial. JMIR Ment Health 2021;8(11):e30590 [FREE Full text] [doi: 10.2196/30590] [Medline: 34813486]
63. Lee J, Lim Y, Graham SJ, Kim G, Wiederhold BK, Wiederhold MD, et al. Nicotine craving and cue exposure therapy by using virtual environments. Cyberpsychol Behav 2004;7(6):705-713. [doi: 10.1089/cpb.2004.7.705] [Medline: 15687806]
64. Rebenitsch L, Owen C. Review on cybersickness in applications and visual displays. Virtual Real 2016;20(2):101-125. [doi: 10.1007/s10055-016-0285-9]
65. McCauley M, Sharkey TJ. Cybersickness: perception of self-motion in virtual environments. Presence: Teleoperators & Virtual Environments 1992 Jan 6;1(3):311-318 [FREE Full text] [doi: 10.1162/pres.1992.1.3.311]
66. Dużmańska N, Strojny P, Strojny A. Can simulator sickness be avoided? A review on temporal aspects of smulator sickness. Front Psychol 2018;9:2132 [FREE Full text] [doi: 10.3389/fpsyg.2018.02132] [Medline: 30459688]
67. Chen P, Jindani F, Perry J, Turner NL. Mindfulness and problem gambling treatment. Asian J of Gambling Issues and Public Health 2014;4(1):2. [doi: 10.1186/2195-3007-4-2]


## Abbreviations

**AR:** augmented reality
**CBT:** cognitive behavioral therapy
**CET:** cue exposure therapy
**ICD-11:** International Classification of Diseases, 11th revision
**JBI:** Joanna Briggs Institute
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
**VR:** virtual reality
**WHO:** World Health Organization
**YIAS:** Young's Internet Addiction Scale



XSL•FO
RenderX

*Edited by M Balcarras; submitted 06.May.2025; peer-reviewed by A Eisingerich; comments to author 11.Sep.2025; revised version received 09.Jan.2026; accepted 12.Jan.2026; published 30.Apr.2026.*

*Please cite as:*
*Chiew FXR, Li GH, Tseung V, Shi J*
*Treating Behavioral Addictions With Augmented Reality and Virtual Reality: Scoping Review*
*Interact J Med Res 2026;15:e77011*
*URL: https://www.i-jmr.org/2026/1/e77011*
*doi:10.2196/77011*
*PMID:42060914*

©Felicia Xin Rou Chiew, Genevieve Huimin Li, Victrine Tseung, Jing Shi. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 30.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Review

# Constructs Influencing Patient Perceptions of Use of AI in Medical Imaging Analysis: Systematic Review

Preksha Machaiya Kuppanda[1], MSc; Monika Janda[1], PhD; Liam J Caffery[1,2], PhD

[1]Frazer Institute, Faculty of Health, Medicine and Behavioural Sciences, The University of Queensland, Brisbane, Queensland, Australia

[2]Centre for Online Health, The University of Queensland, Brisbane, Queensland, Australia

**Corresponding Author:**
Preksha Machaiya Kuppanda, MSc
Frazer Institute
Faculty of Health, Medicine and Behavioural Sciences
The University of Queensland
34 Cornwall St, Woolloongabba QLD
Brisbane, Queensland, 4102
Australia
Phone: 61 734434357
Email: p.kuppanda@uq.edu.au

## Abstract

**Background:** The use of artificial intelligence (AI) in medical imaging has been growing exponentially. Understanding patient perceptions and factors influencing their views of AI is critical to develop adequate strategies to support implementation and acceptance.

**Objective:** This study aims to investigate the constructs that influence patients' perceptions and acceptance of AI's use in the analysis of their medical images to support screening and diagnosis.

**Methods:** A systematic review was conducted to meet the research objective. Relevant articles were found by searching 5 databases. Data were extracted using an iteratively refined framework and synthesized narratively due to heterogeneity in study designs, populations, health care contexts, and outcomes.

**Results:** A total of 59 relevant studies were included in the review. Patient acceptance of AI in medical image analysis emerged from multiple interacting factors. The most consistently reported determinant in 48 studies was that AI implementation should prioritize human-in-the-loop models, positioning AI as supportive tools, working in conjunction with health care providers rather than as an autonomous decision-maker. Other factors identified were performance of the AI, clarity of accountability, trust, and ethical factors. Patients' individual characteristics such as demographics and health history were also noted to influence acceptance indirectly. The review findings were used to draft a conceptual model to draw attention to the complex relationship among the identified factors.

**Conclusions:** This review informed the development of a conceptual model illustrating the complex and interactive factors shaping patient acceptance of AI in medical imaging, which can be tested prospectively in future studies. Our results highlight that patients' likelihood of accepting AI cannot be attributed to a few factors. Instead, promoting acceptance will require a holistic approach where multiple factors are considered simultaneously and adapted for each use case.

***(Interact J Med Res 2026;15:e92969)*** doi:10.2196/92969

**KEYWORDS**

artificial intelligence; patient-centered care; medical imaging; patient engagement; health technology

## Introduction

The use of artificial intelligence (AI) to support medical imaging analysis has been growing exponentially [1,2]. AI tools are increasingly used for a variety of tasks, ranging from automated image acquisition, sophisticated image reconstruction, classification, and diagnosis [3]. Initially, AI tools were primarily used for tasks like image segmentation or abnormality detection, whereas recent applications have advanced to deep learning–driven complex medical image classification based on large datasets, with minimal human involvement [4]. For example, AI applications have shown promising results and



XSL•FO
RenderX

capabilities in the early detection of lung [5], skin [6], and breast cancers [7] from various image types such as computed tomography scans, dermoscopic (magnified) skin images, and mammography scans. In experimental studies, AI's performance has been comparable or surpassed human readers, including experienced clinicians [8-10]. In the future, AI is expected to contribute widely to the imaging process, including image analysis and interpretation, and reduce burden on health care professionals, ultimately contributing to improved patient outcomes [1,11,12].

Although the potential benefits of AI in medical image analysis are widely recognized and its implementation is proceeding rapidly, its application in health care remains largely confined to experimental and research settings, with limited integration into clinical practice in some medical imaging applications [13]. Lack of confidence and skepticism among health care providers and patients toward novel health care technologies have been reported as contributing factors [14,15]. As technology improves, understanding if perceptions and views change is essential to develop adequate strategies to support the implementation and acceptance of AI [16].

Most existing studies and reviews to date have addressed health care professional [17] or multi-stakeholder [18] perceptions and the factors influencing their acceptance of AI in medical image analysis. Research has explored patient perceptions of AI implementation across various medical imaging domains, identifying factors such as trust, human interaction, transparency, accuracy, and individual-perceived benefits and risks associated with AI as key influences [19]. However, despite this progress, there remains a gap in understanding the full range of factors and how they interact to shape patients' perceptions of AI. Recognizing the multitude of factors and how they interact can enable implementation scientists to suggest strategies to make the AI more acceptable to patients. Furthermore, this can guide the development of information leaflets and consent forms as well as implementation strategies that are patient-centered and allow informed consent of AI use across future use cases.

The aim of this review was therefore to identify the factors that impact patients' views of AI use in medical imaging applications and to understand how frequently they are reported, how these variables interact, or how contextual elements such as patient demographics, health issues, or the role of AI in the health care process might shape perceptions. The literature was searched with the aim to (1) identify the factors influencing patients' perceptions of AI in medical image analysis and (2) understand how these factors influence patients' acceptance of AI.

## Methods

### Study Design

We undertook a systematic review of the literature. The SPICE [20] (Setting, Population or perspective, Intervention, Comparison, Evaluation) framework was applied to structure the research question and guide the review process, including the development of the search strategy, establishing the inclusion/exclusion criteria, and selection of relevant articles (Table 1). We followed a 2-phase search strategy (Multimedia Appendix 1), which was developed with the assistance of a university librarian. The protocol of this systematic review was registered with PROSPERO (CRD42024571707) [21]. The finding of this review was reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analysis) guidelines [22].

**Table 1.** SPICE (Setting, Population or perspective, Intervention, Comparison, Evaluation) framework.

| SPICE framework | Question | Study details |
|---|---|---|
| Setting | What is the context for your question? | Medical imaging AI[a] |
| Population/Perspective | What is the perspective of your stakeholders or future users of the service or product? | Patient perspectives on medical imaging AI |
| Intervention | What action is being taken? | Health care practitioner using AI in medical image analysis |
| Comparison | Are there other/alternative actions or outcomes? | Traditional medical image analysis (if applicable) |
| Evaluation | What is the outcome measure? | Factors influencing patient perceptions of AI in medical image analysis either independently or in combination with one another |

[a]AI: artificial intelligence.

### Screening, Selection, Quality Appraisal, and Data Extraction

The final search was conducted in August 2025. Articles from the search results were uploaded to Covidence for title and abstract screening, followed by review of full text articles based on the inclusion criteria (see Table 2). In this review and its inclusion, the term "patient" refers to either participants recruited in a health care setting or individuals from the public who are potential patients or user personas representing patients. The term "AI" refers to both hypothetical and established/applied interventions, incorporated either in the clinic to analyze, interpret, and process medical images and/or support the screening, assessment or diagnosis process. The term "acceptance" refers broadly to patients' willingness for AI to be used in their medical image analysis, commonly assessed through their expectations, preference, or comfort with AI outputs and AI use.

The screening of titles and abstracts was independently conducted by 2 reviewers (PMK and LJC). In the instance of disagreements between the reviewers, conflicts were resolved by discussion and comparison of the article under question



XSL•FO
RenderX

against the predetermined inclusion and exclusion criteria. Full text review, quality appraisal, and data extraction were performed primarily by one reviewer (PMK); a random subset of articles was assessed by a second reviewer (LJC) to ensure accuracy; and any discrepancies in the extracted data were resolved through discussion between the reviewers.

The Joanna Briggs Institute (JBI) critical appraisal checklist [23] was used to assess study quality, with the appropriate tool applied based on each study's design (analytical cross-sectional/qualitative). For mixed methods studies, the analytical cross-sectional and qualitative checklists were combined to evaluate the relevant components. As each checklist varies in length, we calculated the proportion of "yes" and "no" responses relative to the total number of relevant questions, excluding items not applicable to the study to compare study quality. Quality appraisal findings were used to inform interpretation of the results; therefore, no studies were excluded based on the quality scores alone.

The data extraction was conducted using an iteratively developed extraction framework, where the extraction fields were determined inductively as patterns emerged across the included studies. This approach was selected to accommodate the conceptual and methodological diversity of the literature and to ensure the comprehensive capture of patient-reported factors influencing perceptions and acceptance of AI in medical image analysis. Data extraction included the author, year of publication, study location, study type, health care context, AI context, recruitment details, number of participants, and the outcome measures (ie, factors influencing patient perceptions of use of AI in their medical image analysis). In this review, a factor was defined as any reported patient-related perception, attitude, trust, concern, acceptance, or preferences regarding the use of AI in medical image analysis.

**Table 2.** Inclusion and exclusion criteria.

| Category | Inclusion | Exclusion |
|---|---|---|
| General | Peer-reviewed, original research articles in English | Reviews, commentaries, non-English language articles |
| AI[a] intervention | AI that analyzes medical images for diagnosis or screening (applied AI or hypothetical contexts) | AI for nondiagnostic tasks (eg, automated image capture, radiotherapy treatment planning, surgical automation) |
| Purpose/use | Used in clinical care and/or by clinicians to support decision making | Patient-facing AI (eg, direct-to-consumer tools, chatbots, mHealth app-based AI) |
| Patient focus | Examines patient perceptions, attitudes, trust, acceptance, or experience | No patient-reported factors or perceptions. Perceptions of nonpatient groups such as health care providers, health care industry representatives, or medical students |
| Outcomes of interest | Reports factors influencing patient acceptance or experience of AI | Studies that do not identify influencing factors |
| Setting and context | Imaging in clinical screening/diagnosis settings (eg, radiology, dermatology, ophthalmology, dentistry) | Nonclinical settings or unrelated applications |

[a]AI: artificial intelligence.

### Analysis

A narrative synthesis approach was used for data analysis, as the included studies had heterogeneous methods, population groups, health care specialty, and context of AI. Findings were compared across studies and organized into conceptually similar groups, which were summarized and synthesized narratively to address the research question. As no quantitative synthesis of the data was undertaken, the patterns described here should be understood as conceptual.

## Results

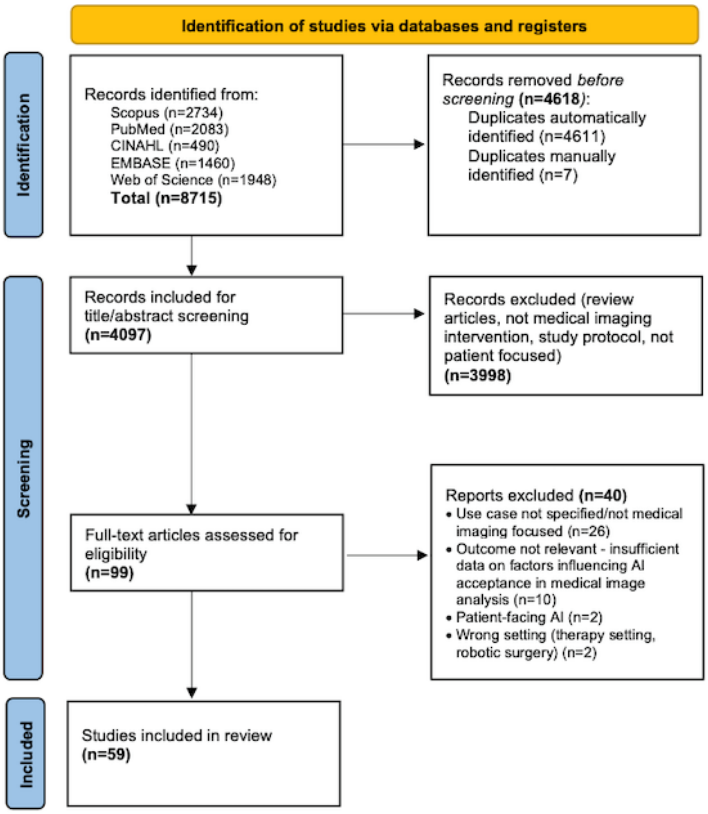
### Screening

In this review, 59 studies [15,19,24-80] were included for data extraction, quality appraisal, and analysis. The results of the screening process are detailed in the PRISMA diagram (Figure 1).



XSL•FO
RenderX

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) flowchart of screening and study selection. AI: artificial intelligence.




## Quality Appraisal

The JBI quality score for the articles ranged from 0.5 to 1 (Table S3 in Multimedia Appendix 1), with higher scores indicating higher quality. Qualitative studies scored highly overall, with lower scores mainly reflecting limited reporting of philosophical positioning. Cross-sectional studies showed greater variability where lower scores were most commonly due to use of nonvalidated questionnaires and limited handling of confounding factors.

## Study Characteristics

The characteristics of the 59 included studies are summarized in Table 3. All included studies were cross-sectional, and 43 studies were quantitative survey–based, with majority (n=20) utilizing de novo, nonvalidated questionnaires, which were informed from literature reviews, existing nonvalidated questionnaires, or expert consultations. Ten included survey studies also adapted the validated “Patients’ Views on the Implementation of AI in Radiology questionnaire” [19]. The qualitative studies mainly used de novo interview guides, with one using prompts adopted from the Theoretical Framework of Acceptability [32] and another mapping its findings to the Theoretical Framework of Acceptability [45]. The included studies were conducted predominantly in Europe (n=34) and focused on radiology-oriented imaging contexts such as mammography (n=13) and general radiology (n=12). The majority of the studies examined hypothetical AI scenarios (n=49).



XSL•FO
RenderX

**Table 3.** Study characteristics (N=59).

| Characteristic | Values, n (%) |
|---|---|
| **Study type** | |
| **Quantitative cross-sectional** | 43 (73) |
| De novo questionnaire, nonvalidated | 20 (34) |
| Adapted from a validated standardized instrument [19] | 10 (17) |
| Other validated standardized instruments (eg, Trust in Medical Technology, Client Satisfaction Questionnaire-8 items) | 3 (5) |
| De novo questionnaire, partially validated | 6 (10) |
| Preference-elicitation instruments (discreet choice/choice-based survey) | 4 (7) |
| **Qualitative cross-sectional** | 13 (22) |
| Qualitative open-ended survey | 1 (2) |
| Focus/dialogue group/engagement workshops | 7 (12) |
| Semistructured interviews | 5 (8) |
| Mixed methods | 3 (5) |
| **Study location** | |
| Europe | 34 (57) |
| Asia | 4 (7) |
| Oceania | 7 (12) |
| Middle East | 3 (5) |
| North America | 7 (12) |
| South America | 1 (2) |
| Africa | 2 (3) |
| Worldwide | 1 (2) |
| **Medical domain** | |
| Mammography/breast imaging | 13 (22) |
| Radiology (general, computed tomography scans) | 12 (20) |
| Dermatology | 9 (15) |
| Ophthalmology | 8 (13) |
| Dental | 5 (8) |
| Radiology (prostate cancer) | 3 (5) |
| Radiology (skeletal) | 4 (7) |
| Cardiology | 1 (2) |
| Imaging (cervical cancer) | 1 (2) |
| Neurosurgery – brain imaging | 1 (2) |
| Gastroenterology | 1 (2) |
| Dermatology + radiology | 1 (2) |
| **AI[a] application context** | |
| Hypothetical AI[b] | 49 (83) |
| Applied AI[c] | 10 (17) |

[a]AI: artificial intelligence.

[b]Hypothetical AI: studies were classified as hypothetical AI when AI or its output was not shown to patients or did not influence clinical care and perceptions were captured through hypothetical scenarios or questions about AI use.

[c]Applied AI: studies were classified as applied AI when AI results were either shown to patients or used by their health care practitioner to inform clinical decisions, allowing patients to experience its presence and report perceptions in the context of actual AI use.



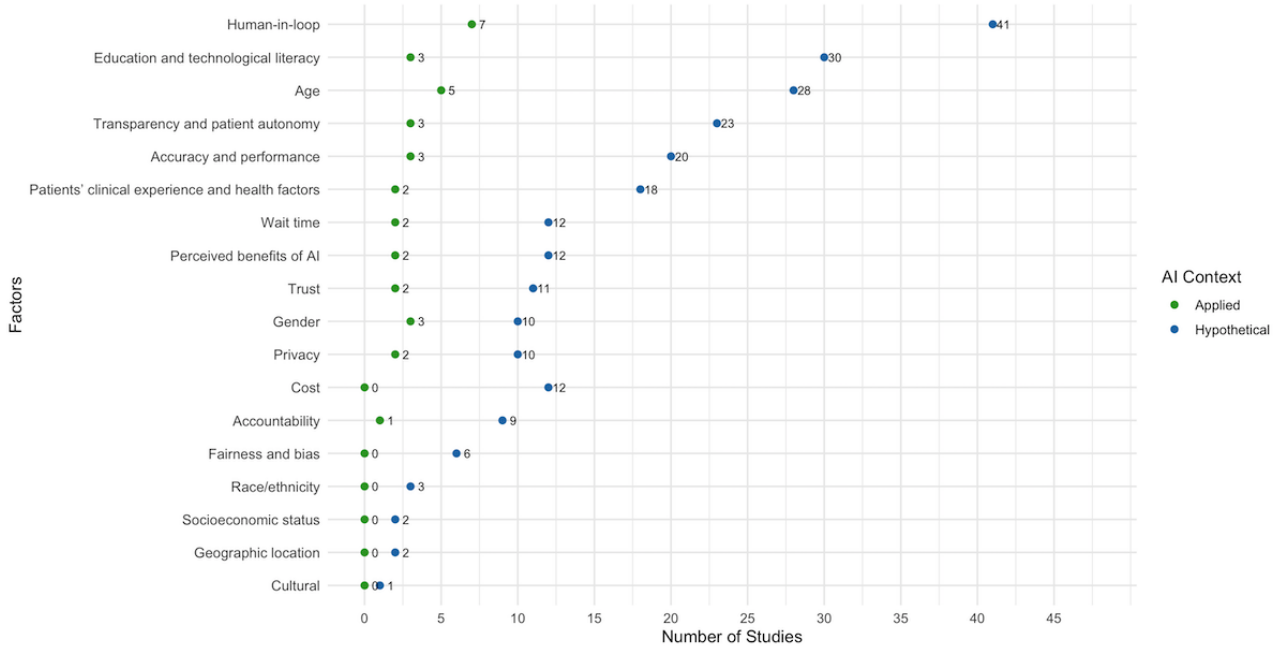
XSL•FO
RenderX

## Results for Aim (a): Factors Influencing Patients' Perceptions of AI in Medical Image Analysis

Eighteen factors that influence patient acceptance and perception of AI in medical image analysis were identified and extracted from the 59 included studies (Table S2 in Multimedia Appendix 1). An overview of the factors identified and the frequency of their reporting is depicted in Figure 2.

Human-in-the-loop was identified as the most frequently reported construct across studies and as a key factor influencing and shaping patient acceptance of AI in their medical image analysis. A sensitivity style examination revealed that for both hypothetical and applied AI contexts individually, the top 5 constructs identified were consistent, although their relative ranking varied slightly in the smaller applied AI sample. This, therefore, suggests that perceptions elicited in hypothetical scenarios broadly reflect those observed in applied clinical settings.

**Figure 2.** Frequency of the extracted factors across studies with hypothetical and applied artificial intelligence (AI) contexts.




## Results for Aim (b): How the Identified Factors Influence Patients' Acceptance of AI

Based on findings from aim (a) and the factors identified, patients' acceptance of AI in medical image analysis appears to arise through a complex interaction among factors, rather than major effects of individual factors.

Figure 3 is a visual representation of the multifaceted relationships between factors and serves as a conceptual model grounded in associations reported across the included studies. Arrows represent influences identified through narrative synthesis and reflect patterns in participants' preferences and perceptions rather than empirically tested causal pathways. The model should be interpreted as an author-synthesized framework illustrating how factors discussed across studies may interact in shaping patient acceptance of AI in medical image analysis and viewed alongside the detailed narrative synthesis in the Multimedia Appendix 1 for full context.

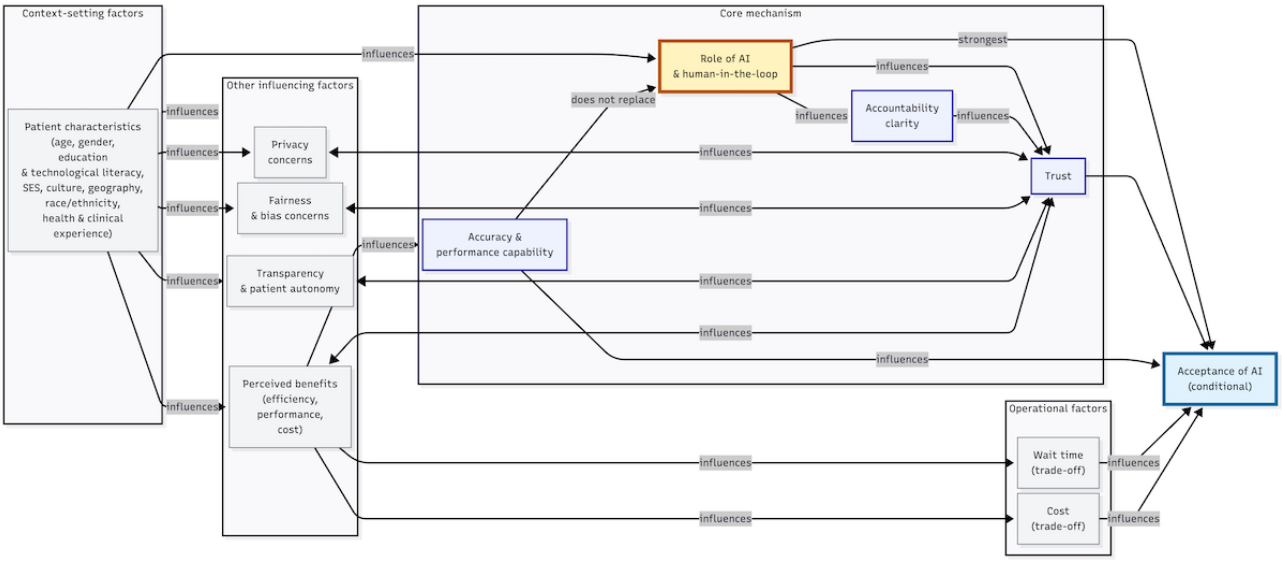
Across multiple studies, human-in-the-loop deployment was reported as a central factor associated with higher acceptance of AI. Accountability was frequently discussed alongside human-in-the-loop deployment, reflecting patients' expectations that clinicians retain responsibility and oversight of AI use. Trust appeared central in how patients evaluated AI, shaping how perceived benefits, risks, and implementation features translated into acceptance. The accuracy of the AI system was commonly reported as a necessary factor for acceptance; however, its impact on acceptance is dependent on human involvement and accountable care models.

Studies reported that the preference and importance placed on the human-in-the-loop model was shaped by patients' individual characteristics such as demographic factors, health issues and history, and sociodemographic characteristics. These factors serve as contextual factors and influence elements such as patients' perceived benefits of AI, their privacy, fairness, and bias concerns, and demand for transparency and autonomy, which in turn shapes trust. Operational factors such as cost and wait time were also noted to influence the acceptance of AI, either as perceived benefits or as trade-offs, with some patients willing to compromise on factors such as human involvement, to achieve improvements in operational conditions or to tolerate less favorable operational conditions in return for retaining human involvement.



XSL•FO
RenderX

**Figure 3.** Synthesized conceptual model of factors influencing patients' acceptance of artificial intelligence (AI) in medical image analysis. SES: socioeconomic status.




## Discussion

Patients' acceptance of AI in medical image analysis was observed to be influenced by several interacting factors. How the AI was applied in the health care setting and human involvement (ie, human-in-the-loop) was most consistently reported to shape acceptance. Hence, high acceptance was reported by studies when AI was positioned as a supportive tool assisting clinicians rather than as functioning autonomously. Other common influential factors included AI accuracy and performance, transparency that AI was used, trust in AI, clarity on who is accountable, and ethical concerns such as privacy and fairness. Sociodemographic characteristics, clinical and health experiences, and operational factors such as waiting time and cost shaped acceptance in more variable and context-dependent ways.

The level of human involvement in how AI is used had the most profound impact on patients' acceptance across studies. These findings are consistent with existing ethical guidelines and empirical studies emphasizing the importance of human oversight and accountability in AI deployment [81-83]. For example, the High-Level Expert Group on Artificial Intelligence listed "human agency and oversight" and "accountability" as a key requirement of a trustworthy AI. Furthermore, Australia's AI ethics principles state that "People responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled." Furthermore, prior work has shown that acceptance of AI-supported services increases when users are informed that humans remain in control, while fear of human replacement can undermine acceptance [84-88]. This pattern is reflected in many emerging AI applications in medical imaging, where systems are typically designed to support clinician decision-making rather than replace it, such as AI tools developed for ultrasound-based disease classification and magnetic resonance imaging lesion detection [89,90].

AI-based image analysis systems are increasingly capable of high diagnostic performance, reflecting the ongoing emphasis among the emerging AI interventions to enhance diagnostic accuracy [91-93]. However, as identified in this review, although the accuracy of AI was frequently prioritized by patients, it did not override the preference for human involvement in the use of AI. Similarly, the emerging AI systems are typically designed to support clinician decision-making rather than operate autonomously, reflecting that these models account for human-in-the-loop requirement. Furthermore, similar to previous research, this review found that transparency and explainability support acceptance primarily by boosting patients' trust rather than through inducing a detailed technical understanding of AI among them [94-96].

Studies reported that factors such as privacy and fairness influenced how patients interpret the benefits and risks of AI, thereby shaping their trust and accordingly acted as either facilitators or barriers of acceptance. Concerns related to privacy, fairness, and bias among patients mirror growing evidence that these concerns can undermine trust in AI-enabled health care [88]. Age, gender, and education of patients were the most commonly studied demographic factors; however, correlations with patients' optimism toward AI were not consistent. The inconsistent influence observed in this review aligns with literature on technology acceptance, which also found that demographic characteristics do not directly influence acceptance. Studies on technology acceptance more broadly not specific to AI found that demographic characteristics such as education shape an individual's technological familiarity, perceived benefits, and expectations of the intervention, which ultimately influences their acceptance [97]. Despite this, acceptance can be high or low within each demographic subgroup and can change over time, as more information about a new technology becomes available and therefore needs to be assessed in future studies.

Patients' individual characteristics like their medical history, experience with clinical contexts, and personal psychological factors were noted to act as context-setting factors, influencing



XSL•FO
RenderX

expectations and sensitivities around AI. Operational factors such as waiting times and costs were identified as trade-offs, with patients potentially willing to trade-off some of their health care professionals' involvement and accept autonomous use of AI in their screening or diagnosis for shorter wait times and reduced costs, or vice versa.

This study has several limitations. First, the conceptual model is grounded in data extracted from the included studies and informed by the authors' interpretation; however, given the complexity of the relationships identified, not all potential interactions could be exhaustively represented. Second, the heterogeneous nature of the included studies and the use of varying survey tools limited direct comparison and quantitative synthesis. Third, certain factors such as cultural influence, socioeconomic status, and race or ethnicity were underreported across studies. This may limit the depth of conclusions that can be drawn in these areas and bias the findings toward populations and health care settings most commonly represented in the literature. This, therefore, may limit the generalizability of conclusions to more diverse contexts. Fourth, variations in terminology and conceptualization of "acceptance" across studies may affect comparability. Fifth, while the search terms and search strategy was developed with adequate guidance to be as exhaustive as possible, there is potential that some relevant publications may have not made into the search results due to variations in the keywords or controlled vocabulary used. Finally, our review focused on peer-reviewed sources reporting original research, including letters or short reports with sufficient methodological detail. Therefore, the findings may emphasize concerns that are more likely to be published, such as privacy, fairness, or specific medical contexts, and may underrepresent perspectives reported in conference proceedings, regional reports, or grey literature. This limitation should be considered when interpreting the results and designing future studies.

Overall, this review suggests that successful implementation of AI in medical image analysis should prioritize human-in-the-loop designs. Policymakers and health care organizations should focus on building governance frameworks that strengthen trust, address privacy and fairness concerns, and ensure that responsibility for AI-assisted decisions remains clearly assigned to human professionals. Furthermore, considering the complex relationship among these influencing factors, with one impacting the other in most instances, patients' choice of accepting AI cannot be attributed to one or a few factors. Instead, promoting acceptance will require a holistic approach where all influencing factors are considered on a case-by-case basis.

Having a standard set of constructs and variables to assess patients' acceptance may be beneficial and enable the development of a structured approach to enhance patient acceptance of AI. The influencing factors identified in this review can form the basis for the development of standard measurement constructs to understand patients' acceptability of AI in medical image analysis for screening and diagnosis. These constructs can be further explored and substantiated by discussions and interviews with patients.

It is important to note that the relationships represented in Figure 3 should not be interpreted as causal pathways. Most included studies were cross-sectional surveys that assessed reported preferences or perceptions, often using hypothetical scenarios. As such, the model represents an interpretive synthesis of reported associations and patterns in participants' reasoning rather than empirically tested mechanisms. The framework is therefore intended to guide future research that could prospectively examine these relationships.

## Acknowledgments

The authors would like to thank the university librarian for their assistance in developing and refining the search strategy for this systematic review.

## Funding

This research is partly funded by the National Health and Medical Research Council (NHMRC) Centre of Research Excellence grant (2006551) and NHMRC Synergy grant (2009923). PMK is supported by a research training program scholarship from the University of Queensland. MJ is funded by an investigator research fellowship from the NHMRC (2034422).

## Data Availability

This systematic review was a secondary study; therefore, data availability is not applicable to this article. Details of the search strategy are included in the Multimedia Appendix 1.

## Authors' Contributions

Conceptualization: PMK, MJ, LJC
Data Curation: PMK
Formal analysis: PMK, LJC
Funding acquisition: MJ
Investigation: PMK, LJC
Methodology: PMK, MJ, LJC



XSL•FO
RenderX

Supervision: LJC
Writing – original draft: PMK
Writing – review and editing: PMK, MJ, LJC
All authors have read and approved the final manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 135 KB - ijmr_v15i1e92969_app1.docx ]

## References


1. Alexander A, Jiang A, Ferreira C, Zurkiya D. An intelligent future for medical imaging: a market outlook on artificial intelligence for medical imaging. J Am Coll Radiol 2020 Jan;17(1 Pt B):165-170 [FREE Full text] [doi: 10.1016/j.jacr.2019.07.019] [Medline: 31918875]
2. Prakash D, Gupta A. Introduction to computer-aided diagnosis (CAD) tools and applications. In: Gupta A, Neelapu BC, Rana SS, editors. Computer-Aided Diagnosis (CAD) Tools and Applications for 3D Medical Imaging, Advances in Computers. Amsterdam: Elsevier; Jan 20, 2025:1-69.
3. Najjar R. Redefining radiology: a review of artificial intelligence integration in medical imaging. Diagnostics (Basel) 2023 Aug 25;13(17):2760 [FREE Full text] [doi: 10.3390/diagnostics13172760] [Medline: 37685300]
4. Savadjiev P, Chong J, Dohan A, Vakalopoulou M, Reinhold C, Paragios N, et al. Demystification of AI-driven medical image interpretation: past, present and future. Eur Radiol 2019 Mar;29(3):1616-1624. [doi: 10.1007/s00330-018-5674-x] [Medline: 30105410]
5. Cellina M, Cacioppa LM, Cè M, Chiarpenello V, Costa M, Vincenzo Z, et al. Artificial intelligence in lung cancer screening: the future is now. Cancers (Basel) 2023 Aug 30;15(17):4344 [FREE Full text] [doi: 10.3390/cancers15174344] [Medline: 37686619]
6. Gohil Z, Desai M. Revolutionizing dermatology: a comprehensive survey of AI-enhanced early skin cancer diagnosis. Arch Computat Methods Eng 2024 Apr 23;31(8):4521-4531 [FREE Full text] [doi: 10.1007/s11831-024-10121-7]
7. Darbandi MR, Darbandi M, Darbandi S, Bado I, Hadizadeh M, Khorram Khorshid HR. Artificial intelligence breakthroughs in pioneering early diagnosis and precision treatment of breast cancer: A multimethod study. Eur J Cancer 2024 Sep;209:114227. [doi: 10.1016/j.ejca.2024.114227] [Medline: 39053289]
8. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health 2019 Oct;1(6):e271-e297 [FREE Full text] [doi: 10.1016/S2589-7500(19)30123-2] [Medline: 33323251]
9. Wu JT, Wong KCL, Gur Y, Ansari N, Karargyris A, Sharma A, et al. Comparison of chest radiograph interpretations by artificial intelligence algorithm vs radiology residents. JAMA Netw Open 2020 Oct 01;3(10):e2022779 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.22779] [Medline: 33034642]
10. Tschandl P, Codella N, Akay BN, Argenziano G, Braun RP, Cabo H, et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. Lancet Oncol 2019 Jul;20(7):938-947 [FREE Full text] [doi: 10.1016/S1470-2045(19)30333-X] [Medline: 31201137]
11. Athavale R, Gutiérrez V, Jha S. AI in medicine: an introduction to the potential benefits and challenges, and why doctors need to be involved. The Obstetric & Gynaecologis 2024 Sep 26;26(4):177-182 [FREE Full text] [doi: 10.1111/tog.12950]
12. Ali O, Abdelbaki W, Shrestha A, Elbasi E, Alryalat M, Dwivedi Y. A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities. Journal of Innovation & Knowledge 2023 Jan;8(1):100333 [FREE Full text] [doi: 10.1016/j.jik.2023.100333]
13. Laï MC, Brian M, Mamzer MF. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. J Transl Med 2020 Jan 09;18(1):14 [FREE Full text] [doi: 10.1186/s12967-019-02204-y] [Medline: 31918710]
14. Tyson A, Pasquini G, Spencer A, Funk C. 60% of Americans would be uncomfortable with provider relying on AI in their own health care. Pew Research Center. 2023 Feb 22. URL: https://www.pewresearch.org/science/2023/02/22/60-of-americans-would-be-uncomfortable-with-provider-relying-on-ai-in-their-own-health-care/ [accessed 2025-12-24]
15. Ibba S, Tancredi C, Fantesini A, Cellina M, Presta R, Montanari R, et al. How do patients perceive the AI-radiologists interaction? Results of a survey on 2119 responders. Eur J Radiol 2023 Aug;165:110917. [doi: 10.1016/j.ejrad.2023.110917] [Medline: 37327548]
16. Shamszare H, Choudhury A. Clinicians' perceptions of artificial intelligence: focus on workload, risk, trust, clinical decision making, and clinical integration. Healthcare (Basel) 2023 Aug 16;11(16):2308 [FREE Full text] [doi: 10.3390/healthcare11162308] [Medline: 37628506]




XSL•FO
RenderX

17. Hua D, Petrina N, Young N, Cho JG, Poon SK. Understanding the factors influencing acceptability of AI in medical imaging domains among healthcare professionals: a scoping review. Artif Intell Med 2024 Jan;147:102698 [FREE Full text] [doi: 10.1016/j.artmed.2023.102698] [Medline: 38184343]
18. Chew HSJ, Achananuparp P. Perceptions and needs of artificial intelligence in health care to increase adoption: scoping review. J Med Internet Res 2022 Jan 14;24(1):e32939 [FREE Full text] [doi: 10.2196/32939] [Medline: 35029538]
19. Ongena YP, Haan M, Yakar D, Kwee TC. Patients' views on the implementation of artificial intelligence in radiology: development and validation of a standardized questionnaire. Eur Radiol 2020 Feb;30(2):1033-1040 [FREE Full text] [doi: 10.1007/s00330-019-06486-0] [Medline: 31705254]
20. Booth A. Clear and present questions: formulating questions for evidence based practice. Library hi tech 2006:355-368 [FREE Full text] [doi: 10.1108/07378830610692127]
21. Kuppanda PM, Caffery LJ. Constructs influencing patient experience or perception of use of artificial intelligence (AI) in medical imaging diagnostics? A systematic review. PROSPERO. 2024 Aug 03. URL: https://www.crd.york.ac.uk/PROSPERO/view/CRD42024571707 [accessed 2024-08-31]
22. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
23. Critical appraisal tools. Joanna Briggs Institute. URL: https://jbi.global/critical-appraisal-tools [accessed 2025-08-31]
24. Xuereb F, Portelli DJL. The knowledge and perception of patients in Malta towards artificial intelligence in medical imaging. J Med Imaging Radiat Sci 2024 Dec;55(4):101743 [FREE Full text] [doi: 10.1016/j.jmir.2024.101743] [Medline: 39317135]
25. Fransen SJ, Kwee TC, Rouw D, Roest C, van Lohuizen QY, Simonis FFJ, et al. Patient perspectives on the use of artificial intelligence in prostate cancer diagnosis on MRI. Eur Radiol 2025 Feb;35(2):769-775. [doi: 10.1007/s00330-024-11012-y] [Medline: 39143247]
26. Popic D, Marinovich ML, Houssami N, Hall J, Carter SM. How should artificial intelligence be used in breast screening? Women's reasoning about workflow options. PLoS One 2025;20(5):e0323528 [FREE Full text] [doi: 10.1371/journal.pone.0323528] [Medline: 40446203]
27. Goessinger EV, Niederfeilner J, Cerminara S, Maul J, Kostner L, Kunz M, et al. Patient and dermatologists' perspectives on augmented intelligence for melanoma screening: A prospective study. J Eur Acad Dermatol Venereol 2024 Dec;38(12):2240-2249. [doi: 10.1111/jdv.19905] [Medline: 38411348]
28. Lee L, Salami RK, Martin H, Shantharam L, Thomas K, Ashworth E, et al. "How I would like AI used for my imaging": children and young persons' perspectives. Eur Radiol 2024 Dec;34(12):7751-7764. [doi: 10.1007/s00330-024-10839-9] [Medline: 38900281]
29. Rodler S, Kopliku R, Ulrich D, Kaltenhauser A, Casuscelli J, Eismann L, et al. Patients' trust in artificial intelligence-based decision-making for localized prostate cancer: results from a prospective trial. Eur Urol Focus 2024 Jul;10(4):654-661. [doi: 10.1016/j.euf.2023.10.020] [Medline: 37923632]
30. Clements W, Thong L, Zia A, Moriarty H, Goh G. A prospective study assessing patient perception of the use of artificial intelligence in radiology. APJHM 2022 Apr 07:1 [FREE Full text] [doi: 10.24083/apjhm.v17i1.861]
31. Fink C, Uhlmann L, Hofmann M, Forschner A, Eigentler T, Garbe C, et al. Patient acceptance and trust in automated computer-assisted diagnosis of melanoma with dermatofluoroscopy. J Dtsch Dermatol Ges 2018 Jul;16(7):854-859. [doi: 10.1111/ddg.13562] [Medline: 29927518]
32. Gatting L, Ahmed S, Meccheri P, Newlands R, Kehagia AA, Waller J. Acceptability of artificial intelligence in breast screening: focus groups with the screening-eligible population in England. BMJ Public Health 2024 Dec;2(2):e000892. [doi: 10.1136/bmjph-2024-000892] [Medline: 40018529]
33. Wahlich C, Chandrasekaran L, Chaudhry UAR, Willis K, Chambers R, Bolter L, et al. Patient and practitioner perceptions around use of artificial intelligence within the English NHS diabetic eye screening programme. Diabetes Res Clin Pract 2025 Jan;219:111964. [doi: 10.1016/j.diabres.2024.111964] [Medline: 39709112]
34. Malerbi FK, Mezzomo Ventura B, Fischer M, Penha FM. Patients perceptions of artificial intelligence in a deep learning-assisted diabetic retinopathy screening event: a real-world assessment. J Diabetes Sci Technol 2024 May;18(3):750-751. [doi: 10.1177/19322968241234378] [Medline: 38404014]
35. Palmisciano P, Jamjoom AAB, Taylor D, Stoyanov D, Marcus HJ. Attitudes of patients and their relatives toward artificial intelligence in neurosurgery. World Neurosurg 2020 Jun;138:e627-e633. [doi: 10.1016/j.wneu.2020.03.029] [Medline: 32179185]
36. Viberg Johansson J, Dembrower K, Strand F, Grauman A. Women's perceptions and attitudes towards the use of AI in mammography in Sweden: a qualitative interview study. BMJ Open 2024 Feb 14;14(2):e084014 [FREE Full text] [doi: 10.1136/bmjopen-2024-084014] [Medline: 38355190]
37. Lennox-Chhugani N, Chen Y, Pearson V, Trzcinski B, James J. Women's attitudes to the use of AI image readers: a case study from a national breast screening programme. BMJ Health Care Inform 2021 Mar;28(1):e100293 [FREE Full text] [doi: 10.1136/bmjhci-2020-100293] [Medline: 33795236]




XSL•FO
RenderX

38. Haan M, Ongena YP, Hommes S, Kwee TC, Yakar D. A qualitative study to understand patient perspective on the use of artificial intelligence in radiology. J Am Coll Radiol 2019 Oct;16(10):1416-1419. [doi: 10.1016/j.jacr.2018.12.043] [Medline: 30878311]
39. El-Sayed MZ, Rawashdeh M, Moossa A, Atfah M, Prajna B, Ali M. Patient perspectives on AI in radiology: insights from the United Arab Emirates. Clin Imaging 2025 Sep;125:110543. [doi: 10.1016/j.clinimag.2025.110543] [Medline: 40513450]
40. Bahakeem BH, Alobaidi SF, Alzahrani AS, Alhasawi R, Alzahrani A, Alqahtani W, et al. The general population's perspectives on implementation of artificial intelligence in radiology in the western region of Saudi Arabia. Cureus 2023 Apr;15(4):e37391 [FREE Full text] [doi: 10.7759/cureus.37391] [Medline: 37182053]
41. Pelayo C, Hoang J, Mora Pinzón M, Lock LJ, Fowlkes C, Stevens CL, et al. Perspectives of Latinx patients with diabetes on teleophthalmology, artificial intelligence-based image interpretation, and virtual care: a qualitative study. Telemed Rep 2023;4(1):317-326 [FREE Full text] [doi: 10.1089/tmr.2023.0045] [Medline: 37908628]
42. Baghdadi LR, Mobeirek AA, Alhudaithi DR, Albenmousa FA, Alhadlaq LS, Alaql MS, et al. Patients' attitudes toward the use of artificial intelligence as a diagnostic tool in radiology in Saudi Arabia: cross-sectional study. JMIR Hum Factors 2024 Aug 07;11:e53108. [doi: 10.2196/53108] [Medline: 39110973]
43. Nelson CA, Pérez-Chada LM, Creadore A, Li SJ, Lo K, Manjaly P, et al. Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. JAMA Dermatol 2020 May 01;156(5):501-512 [FREE Full text] [doi: 10.1001/jamadermatol.2019.5014] [Medline: 32159733]
44. Carter SM, Carolan L, Saint James Aquino Y, Frazer H, Rogers WA, Hall J, et al. Australian women's judgements about using artificial intelligence to read mammograms in breast cancer screening. Digit Health 2023;9:20552076231191057 [FREE Full text] [doi: 10.1177/20552076231191057] [Medline: 37559826]
45. Manning F, Mahmoud A, Meertens R. Understanding patient views and acceptability of predictive software in osteoporosis identification. Radiography (Lond) 2023 Oct;29(6):1046-1053 [FREE Full text] [doi: 10.1016/j.radi.2023.08.011] [Medline: 37734275]
46. Frühauf J, Leinweber B, Fink-Puches R, Ahlgrimm-Siess V, Richtig E, Wolf IH, et al. Patient acceptance and diagnostic utility of automated digital image analysis of pigmented skin lesions. J Eur Acad Dermatol Venereol 2012 Mar;26(3):368-372. [doi: 10.1111/j.1468-3083.2011.04081.x] [Medline: 21504486]
47. Ongena YP, Yakar D, Haan M, Kwee TC. Artificial intelligence in screening mammography: a population survey of women's preferences. J Am Coll Radiol 2021 Jan;18(1 Pt A):79-86 [FREE Full text] [doi: 10.1016/j.jacr.2020.09.042] [Medline: 33058789]
48. Miró Catalina Q, Femenia J, Fuster-Casanovas A, Marin-Gomez FX, Escalé-Besa A, Solé-Casals J, et al. Knowledge and perception of the use of AI and its implementation in the field of radiology: cross-sectional study. J Med Internet Res 2023 Oct 13;25:e50728 [FREE Full text] [doi: 10.2196/50728] [Medline: 37831495]
49. Baillie L, Stewart-Lord A, Thomas N, Frings D. Patients', clinicians' and developers' perspectives and experiences of artificial intelligence in cardiac healthcare: a qualitative study. Digit Health 2025;11:20552076251328578 [FREE Full text] [doi: 10.1177/20552076251328578] [Medline: 40534891]
50. Holen S, Martiniussen MA, Bergan MB, Moshina N, Hovda T, Hofvind S. Women's attitudes and perspectives on the use of artificial intelligence in the assessment of screening mammograms. Eur J Radiol 2024 Jun;175:111431. [doi: 10.1016/j.ejrad.2024.111431] [Medline: 38520804]
51. Arora PC, Sandhu KK, Arora A, Gupta A, Waghmare M, Rampal V. Acceptability of artificial intelligence in dental radiology among patients in India: are we ready for this revolution? Oral Radiol 2025 Jan;41(1):69-77. [doi: 10.1007/s11282-024-00777-z] [Medline: 39384683]
52. Yap A, Wilkinson B, Chen E, Han L, Vaghefi E, Galloway C, et al. Patients perceptions of artificial intelligence in diabetic eye screening. Asia Pac J Ophthalmol (Phila) 2022 May 01;11(3):287-293 [FREE Full text] [doi: 10.1097/APO.0000000000000525] [Medline: 35772087]
53. Foresman G, Biro J, Tran A, MacRae K, Kazi S, Schubel L, et al. Patient perspectives on artificial intelligence in health care: focus group study for diagnostic communication and tool implementation. J Particip Med 2025 Jul 24;17:e69564 [FREE Full text] [doi: 10.2196/69564] [Medline: 40705399]
54. Jonmarker O, Strand F, Brandberg Y, Lindholm P. The future of breast cancer screening: what do participants in a breast cancer screening program think about automation using artificial intelligence? Acta Radiol Open 2019 Dec;8(12):2058460119880315 [FREE Full text] [doi: 10.1177/2058460119880315] [Medline: 31839989]
55. Agarwal G, Salami RK, Lee L, Martin H, Shantharam L, Thomas K, et al. Parental and carer views on the use of AI in imaging for children: a national survey. Insights Imaging 2025 Aug 09;16(1):172. [doi: 10.1186/s13244-025-02021-6] [Medline: 40782183]
56. Ozcan BB, Dogan BE, Xi Y, Knippa EE. Patient perception of artificial intelligence use in interpretation of screening mammograms: a survey study. Radiol Imaging Cancer 2025 May;7(3):e240290. [doi: 10.1148/rycan.240290] [Medline: 40249272]
57. Kosan E, Krois J, Wingenfeld K, Deuter CE, Gaudin R, Schwendicke F. Patients' perspectives on artificial intelligence in dentistry: a controlled study. J Clin Med 2022 Apr 12;11(8):2143 [FREE Full text] [doi: 10.3390/jcm11082143] [Medline: 35456236]




XSL•FO
RenderX

58. Jagemann I, Wensing O, Stegemann M, Hirschfeld G. Acceptance of medical artificial intelligence in skin cancer screening: choice-based conjoint survey. JMIR Form Res 2024 Jan 12;8:e46402 [FREE Full text] [doi: 10.2196/46402] [Medline: 38214959]
59. Kawsar A, Hussain K, Kalsi D, Kemos P, Marsden H, Thomas L. Patient perspectives of artificial intelligence as a medical device in a skin cancer pathway. Front Med (Lausanne) 2023;10:1259595 [FREE Full text] [doi: 10.3389/fmed.2023.1259595] [Medline: 38046409]
60. Pesapane F, Rotili A, Valconi E, Agazzi GM, Montesano M, Penco S, et al. Women's perceptions and attitudes to the use of AI in breast cancer screening: a survey in a cancer referral centre. Br J Radiol 2023 Jan 01;96(1141):20220569. [doi: 10.1259/bjr.20220569] [Medline: 36314388]
61. Woode ME, De Silva Perera U, Degeling C, Aquino YSJ, Houssami N, Carter SM, et al. Preferences for the use of artificial intelligence for breast cancer screening in Australia: a discrete choice experiment. Patient 2025 Sep;18(5):495-510. [doi: 10.1007/s40271-025-00742-w] [Medline: 40347323]
62. McGhee KN, Barrett DJ, Safarini O, Elkassem AA, Eddins JT, Smith AD, et al. Patient preferences for artificial intelligence in medical imaging: a single-center cross-sectional survey. J Imaging Inform Med 2026 Apr;39(2):1100-1112. [doi: 10.1007/s10278-025-01629-w] [Medline: 40775178]
63. Pearce A, Carter S, Frazer HM, Houssami N, Macheras-Magias M, Webb G, et al. Implementing artificial intelligence in breast cancer screening: women's preferences. Cancer 2025 May 01;131(9):e35859 [FREE Full text] [doi: 10.1002/cncr.35859] [Medline: 40262029]
64. Lim K, Neal-Smith G, Mitchell C, Xerri J, Chuanromanee P. Perceptions of the use of artificial intelligence in the diagnosis of skin cancer: an outpatient survey. Clin Exp Dermatol 2022 Mar;47(3):542-546. [doi: 10.1111/ced.14969] [Medline: 34610153]
65. Tirapelli C, Gaêta-Araujo H, Costa ED, Scarfe WC, Oliveira-Santos C, Fischer KM, et al. Patient perceptions of artificial intelligence in dental imaging diagnostics: a multicentre survey. Dentomaxillofac Radiol 2025 Sep 01;54(6):427-436. [doi: 10.1093/dmfr/twaf018] [Medline: 40080713]
66. Lysø EH, Hesjedal MB, Skolbekken J, Solbjør M. Men's sociotechnical imaginaries of artificial intelligence for prostate cancer diagnostics - a focus group study. Soc Sci Med 2024 Apr;347:116771 [FREE Full text] [doi: 10.1016/j.socscimed.2024.116771] [Medline: 38537333]
67. Adams SJ, Tang R, Babyn P. Patient perspectives and priorities regarding artificial intelligence in radiology: opportunities for patient-centered radiology. J Am Coll Radiol 2020 Aug;17(8):1034-1036. [doi: 10.1016/j.jacr.2020.01.007] [Medline: 32068006]
68. Jutzi TB, Krieghoff-Henning EI, Holland-Letz T, Utikal JS, Hauschild A, Schadendorf D, et al. Artificial intelligence in skin cancer diagnostics: the patients' perspective. Front Med (Lausanne) 2020;7:233 [FREE Full text] [doi: 10.3389/fmed.2020.00233] [Medline: 32671078]
69. Whitestone N, Nkurikiye J, Patnaik JL, Jaccard N, Lanouette G, Cherwek DH, et al. Feasibility and acceptance of artificial intelligence-based diabetic retinopathy screening in Rwanda. Br J Ophthalmol 2024 May 21;108(6):840-845. [doi: 10.1136/bjo-2022-322683] [Medline: 37541766]
70. York T, Jenney H, Jones G. Clinician and computer: a study on patient perceptions of artificial intelligence in skeletal radiography. BMJ Health Care Inform 2020 Nov;27(3):e100233 [FREE Full text] [doi: 10.1136/bmjhci-2020-100233] [Medline: 33187956]
71. Ye T, Xue J, He M, Gu J, Lin H, Xu B, et al. Psychosocial factors affecting artificial intelligence adoption in health care in China: cross-sectional study. J Med Internet Res 2019 Oct 17;21(10):e14316 [FREE Full text] [doi: 10.2196/14316] [Medline: 31625950]
72. Yakar D, Ongena YP, Kwee TC, Haan M. Do people favor artificial intelligence over physicians? a survey among the general population and their view on artificial intelligence in medicine. Value Health 2022 Mar;25(3):374-381. [doi: 10.1016/j.jval.2021.09.004] [Medline: 35227448]
73. Shah P, Mishra D, Shanmugam M, Vighnesh MJ, Jayaraj H. Acceptability of artificial intelligence-based retina screening in general population. Indian J Ophthalmol 2022 Apr;70(4):1140-1144 [FREE Full text] [doi: 10.4103/ijo.IJO_1840_21] [Medline: 35326001]
74. Sachdeva M, Datchoua AM, Yakam VF, Kenfack B, Jonnalagedda-Cattin M, Thiran JP, et al. Acceptability of artificial intelligence for cervical cancer screening in Dschang, Cameroon: a qualitative study on patient perspectives. Reprod Health 2024 Jun 28;21(1):92 [FREE Full text] [doi: 10.1186/s12978-024-01828-8] [Medline: 38937771]
75. Schmidt KA, Sood S, Dilmaghani S, Leggett C, Dierkhising R, Goyal M, et al. Understanding patients' current acceptability of artificial intelligence during colonoscopy for polyp detection: a single-center study. Techniques and Innovations in Gastrointestinal Endoscopy 2025;27(2):250905. [doi: 10.1016/j.tige.2024.250905]
76. Keel S, Lee PY, Scheetz J, Li Z, Kotowicz MA, MacIsaac RJ, et al. Feasibility and patient acceptability of a novel artificial intelligence-based screening model for diabetic retinopathy at endocrinology outpatient services: a pilot study. Sci Rep 2018 Mar 12;8(1):4330. [doi: 10.1038/s41598-018-22612-2] [Medline: 29531299]




XSL•FO
RenderX

77. Haggenmüller S, Maron RC, Hekler A, Krieghoff-Henning E, Utikal JS, Gaiser M, Collaborators. Patients' and dermatologists' preferences in artificial intelligence-driven skin cancer diagnostics: A prospective multicentric survey study. J Am Acad Dermatol 2024 Aug;91(2):366-370 [FREE Full text] [doi: 10.1016/j.jaad.2024.04.033] [Medline: 38670313]
78. Dontchos BN, Dodelzon K, Bhole S, Edmonds CE, Mullen LA, Parikh JR, et al. Opinions and preferences regarding artificial intelligence use in health care delivery: results from a national multisite survey of breast imaging patients. J Am Coll Radiol 2025 Sep;22(9):1032-1040. [doi: 10.1016/j.jacr.2025.05.001] [Medline: 40339678]
79. Ayad N, Schwendicke F, Krois J, van den Bosch S, Bergé S, Bohner L, et al. Patients' perspectives on the use of artificial intelligence in dentistry: a regional survey. Head Face Med 2023 Jun 22;19(1):23 [FREE Full text] [doi: 10.1186/s13005-023-00368-z] [Medline: 37349791]
80. Bahadir HS, Keskin NB, Çakmak EŞK, Güneç G, Cesur Aydin K, Peker F. Patients' attitudes toward artificial intelligence in dentistry and their trust in dentists. Oral Radiol 2025 Jan 09;41(1):52-59. [doi: 10.1007/s11282-024-00775-1] [Medline: 39379636]
81. Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. EU-Commission. URL: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment [accessed 2025-11-01]
82. Australia's AI ethics principles. Australian Government, Department of Industry, Science and Resources. 2019 Nov 07. URL: https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-principles [accessed 2025-11-06]
83. Ploug T, Sundby A, Moeslund TB, Holm S. Population preferences for performance and explainability of artificial intelligence in health care: choice-based conjoint survey. J Med Internet Res 2021 Dec 13;23(12):e26611-e26626 [FREE Full text] [doi: 10.2196/26611] [Medline: 34898454]
84. Aoki N. The importance of the assurance that "humans are still in the decision loop" for public trust in artificial intelligence: Evidence from an online experiment. Computers in Human Behavior 2021 Jan;114:106572 [FREE Full text] [doi: 10.1016/j.chb.2020.106572]
85. Tran VT, Riveros C, Ravaud P. Patients' views of wearable devices and AI in healthcare: findings from the ComPaRe e-cohort. NPJ Digit Med 2019;2:53 [FREE Full text] [doi: 10.1038/s41746-019-0132-y] [Medline: 31304399]
86. Gaube S, Biebl I, Engelmann MKM, Kleine AK, Lermer E. Comparing preferences for skin cancer screening: AI-enabled app vs dermatologist. Soc Sci Med 2024 May;349:116871 [FREE Full text] [doi: 10.1016/j.socscimed.2024.116871] [Medline: 38640741]
87. Li Y, Wu B, Huang Y, Luan S. Developing trustworthy artificial intelligence: insights from research on interpersonal, human-automation, and human-AI trust. Front Psychol 2024;15:1382693 [FREE Full text] [doi: 10.3389/fpsyg.2024.1382693] [Medline: 38694439]
88. Gillespie N, Lockey S, Curtis C, Pool J, Akbari A. Trust in artificial intelligence: a global study. The University of Queensland and KPMG Australia. URL: https://ai.uq.edu.au/project/trust-artificial-intelligence-global-study [accessed 2025-09-16]
89. Wu M, Yan C, Sen G. Computer-aided diagnosis of hepatic cystic echinococcosis based on deep transfer learning features from ultrasound images. Sci Rep 2025 Jan 03;15(1):607 [FREE Full text] [doi: 10.1038/s41598-024-85004-9] [Medline: 39753933]
90. Guo T, Luan J, Gao J, Liu B, Shen T, Yu H, et al. Computer-aided diagnosis of pituitary microadenoma on dynamic contrast-enhanced MRI based on spatio-temporal features. Expert Systems with Applications 2025 Jan;260:125414 [FREE Full text] [doi: 10.1016/j.eswa.2024.125414]
91. Chen Y, Feng L, Zheng C, Zhou T, Liu L, Liu P, et al. LDANet: Automatic lung parenchyma segmentation from CT images. Comput Biol Med 2023 Mar;155:106659. [doi: 10.1016/j.compbiomed.2023.106659] [Medline: 36791550]
92. Liang X, Chen S, Shu L, Wang D, Chen Q, Cao Y, et al. Enhancing polyp detection in endoscopy with cross-channel self-attention fusion. Smart Health 2025 Jun;36:100578 [FREE Full text] [doi: 10.1016/j.smhl.2025.100578]
93. He Y, Chen Q, Xiong Z, Liang X, Cao Y, Liu B. One-stage framework for thyroid nodule detection with mixup and negative sample utilization. 2025 Aug 18 Presented at: 2025 IEEE International Conference on Image Processing (ICIP); September 14-17, 2025; Anchorage, AK, USA. [doi: 10.1109/ICIP55913.2025.11084660]
94. Ploug T, Holm S. Right to contest AI diagnostics. In: Lidströmer N, Ashrafian H, editors. Artificial Intelligence in Medicine. Switzerland: Springer; Feb 18, 2022:227-238.
95. Beltrami EJ, Brown AC, Salmon PJM, Leffell DJ, Ko JM, Grant-Kels JM. Artificial intelligence in the detection of skin cancer. J Am Acad Dermatol 2022 Dec;87(6):1336-1342 [FREE Full text] [doi: 10.1016/j.jaad.2022.08.028] [Medline: 35998842]
96. Phung M, Muralidharan V, Rotemberg V, Novoa RA, Chiou AS, Sadée CY, et al. Best practices for clinical skin image acquisition in translational artificial intelligence research. J Invest Dermatol 2023 Jul;143(7):1127-1132 [FREE Full text] [doi: 10.1016/j.jid.2023.02.035] [Medline: 37353282]
97. Porter C, Donthu N. Using the technology acceptance model to explain how attitudes determine internet usage: the role of perceived access barriers and demographics. Journal of Business Research 2006 Sep;59(9):999-1007 [FREE Full text] [doi: 10.1016/j.jbusres.2006.06.003]




XSL•FO
RenderX

## Abbreviations

**AI:** artificial intelligence
**JBI:** Joanna Briggs Institute
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-analysis
**SPICE:** Setting, Population or perspective, Intervention, Comparison, Evaluation

*Edited by S Consoli; submitted 05.Feb.2026; peer-reviewed by X Liang, J Zhang, H Chen; comments to author 10.Mar.2026; accepted 20.Apr.2026; published 01.Jun.2026.*

*Please cite as:*
*Kuppanda PM, Janda M, Caffery LJ*
*Constructs Influencing Patient Perceptions of Use of AI in Medical Imaging Analysis: Systematic Review*
*Interact J Med Res 2026;15:e92969*
*URL: https://www.i-jmr.org/2026/1/e92969*
*doi:10.2196/92969*
*PMID:*

©Preksha Machaiya Kuppanda, Monika Janda, Liam J Caffery. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 01.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# An Introduction to AI for Clinicians: Tutorial

Stephen B Lee[1], MS, MD; Alexis B Carter[2], MD; Muhammad Hamis Haider[3], PhD; Seok-Bum Ko[3], PhD

[1]Division of Infectious Diseases, University of Saskatchewan, 1440-14th Avenue, Regina General Hospital, 2nd Floor Medical Office Wing, ID Clinic, Regina, SK, Canada

[2]Department of Pathology and Laboratory Medicine, Emory University, Atlanta, GA, United States

[3]Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK, Canada

**Corresponding Author:**
Stephen B Lee, MS, MD
Division of Infectious Diseases, University of Saskatchewan, 1440-14th Avenue, Regina General Hospital, 2nd Floor Medical Office Wing, ID Clinic, Regina, SK, Canada

## Abstract

Artificial intelligence (AI) is already fundamentally changing society, with medicine being no exception. AI will impact how we practice, how hospitals operate, and even the practice of medicine itself. The use of AI-based products has already begun, with examples including AI scribes and large language models such as ChatGPT. Work is ongoing to produce models that have specific functions within medicine, such as kidney injury prediction. However, transformative foundational work, such as AlphaFold (for protein structure prediction), also promises to completely change the way we approach medicine. Therefore, clinicians must develop a clear understanding of AI, not as an optional skill, but as a core competency of modern medical practice. This paper serves as a tutorial to guide medical professionals through the basic principles of AI. It will teach clinicians how to build a mental scaffold to understand and springboard into AI. The core parts of this paper are organized in steps, with additional relevant topics addressed in modules at the end of the paper. The core steps are meant to be read sequentially. To prepare the reader for the rest of the paper, this tutorial will first introduce what AI is and then cover some basic definitions needed to understand other concepts. The reader will then be ready to understand what deep learning is and the difference between supervised and unsupervised learning. Finally, the reader will go through how deep learning models learn. Separate modules on safety and clinical applications are also included. This tutorial is relevant to clinicians at all levels but may be particularly useful for practicing clinicians who are encountering AI tools integrated into their practices without previous formal education in the field. Users of this tutorial can refer to specific sections or read the entire paper.

***(Interact J Med Res 2026;15:e85266)*** doi:10.2196/85266

**KEYWORDS**

medical education; machine learning; artificial intelligence; AI; deep learning

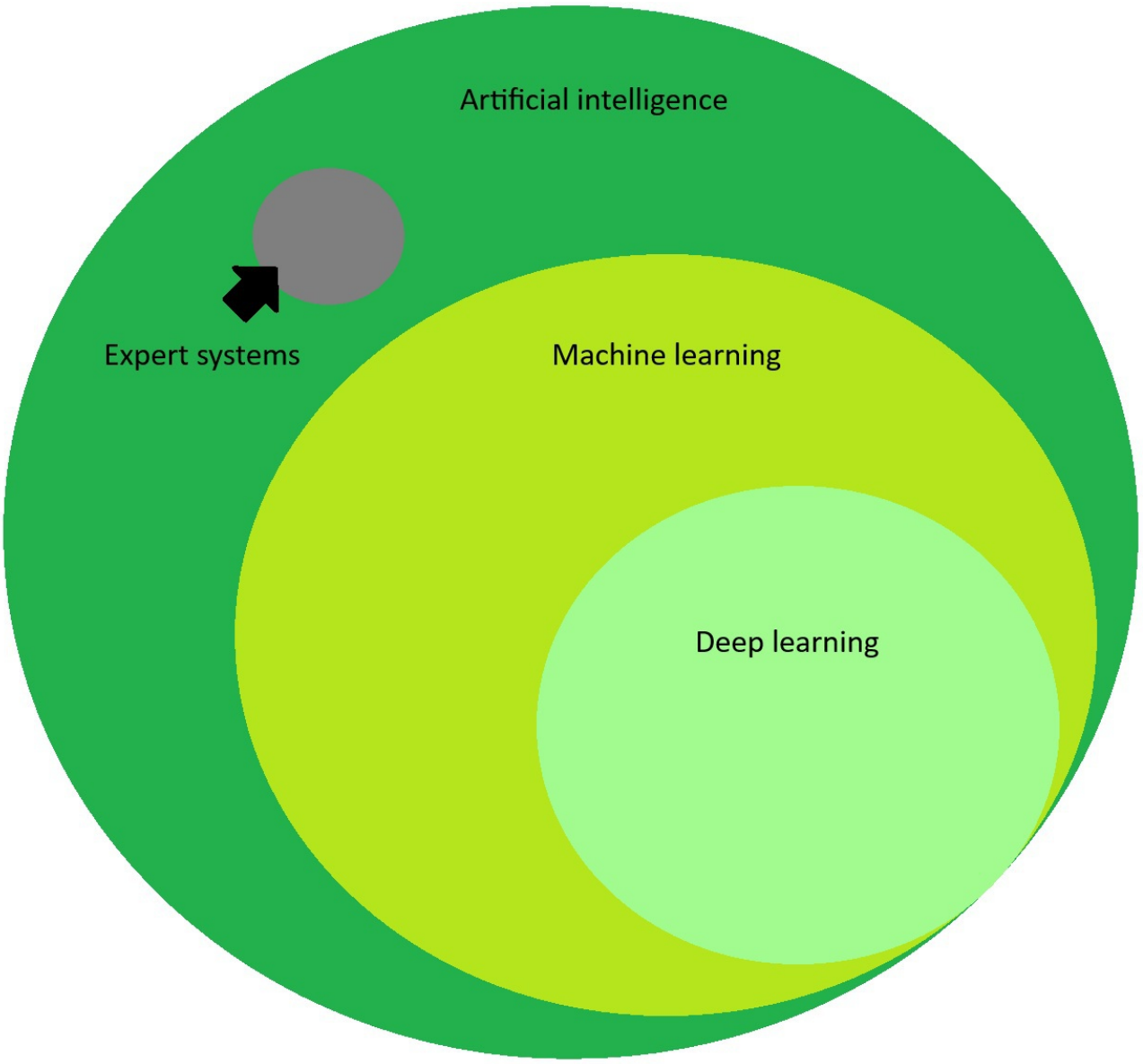
## Step 1: Understanding What Artificial Intelligence Is

Before delving into the concepts underlying artificial intelligence (AI), it is important to understand what AI means. AI is a broad term that generally refers to computer systems that can perform complex tasks historically associated with humans, such as human learning, comprehension, problem-solving, decision-making, creativity, and autonomy [1]. In computer science, AI algorithms encompass a variety of methodologies. Most AI in the modern era is machine learning (ML) and even more specifically, deep learning (DL; Figure 1). While all ML is considered AI, not all AI is ML.



XSL•FO
RenderX

**Figure 1.** Venn diagram of artificial intelligence terminology.




Expert systems are an older paradigm of AI in which a subject matter expert's knowledge is hardcoded into intricate rule-based algorithms to simulate human decision-making (eg, "if x condition, then y result"). Well-known examples include the MYCIN system, designed by Edward Shortliffe in 1974 to predict antibiotic choice [2], and traditional chess AI systems. Even famous examples such as Deep Blue (IBM Corp) relied on hardcoded knowledge [3]. Expert systems and these hardcoded methodologies have been largely abandoned in the modern era due to the incredible effort required to develop them and the brittleness and inflexibility of the resultant systems. Limited examples of expert systems also exist in health care, such as older sepsis and drug interaction clinical decision support systems. AI is primarily ML currently. Rather than relying on hardcoding from experts, ML algorithms are trained to learn relationships and patterns in data. The terms DL and ML are often used interchangeably; however, it is worthwhile mentioning that DL is really a subset of ML. There are many types of ML that are not DL (eg, k-means clustering and decision trees). DL is discussed in depth in a following section.

## Step 2: Understanding Basic Definitions

Some level-setting is required to understand ML algorithms and models. ML *algorithms* are tools (eg, logistic regression) that are trained on data to create an ML *model*. ML models that are used for classification are frequently referred to as *classifiers*. The quality of the data used to train an algorithm is critical to the performance of the model. Each record or event in the dataset is referred to as an *instance*. Each aspect of an instance (eg, color, duration, and test result) used to train a model is known as a *feature*. In a simple set of data, instances would be rows in your data spreadsheet, while features would be the column headers. Many ML datasets can have thousands of instances and hundreds of features.

A *label* is the information on a particular feature for a certain instance in the dataset (eg, "red" for a feature of "color," "30 minutes" for a feature of "duration," and "34 mg/dL" for a feature of "glucose level"). A label generally refers to information that has been applied by a human or another



XSL•FO
RenderX

algorithm based on manual or traditional analysis of the data and represents a classification, categorization, ranking, or answer to a question. For example, in a dataset of inpatients hospitalized for at least 30 days with a feature of "patient developed a hospital-acquired infection during admission," the label would be "positive" if an infection was acquired and "negative" if none occurred, bearing in mind that the application of positive or negative would have to be done by a human who was analyzing the patient's data.

Of note is that data that are not manually labeled are simply referred to as data. One may also choose to use only a subset of the data in a dataset for training.

## Step 3: Differentiating Between Supervised and Unsupervised ML

A core initial concept is understanding the difference between 2 major categories of ML. In general, ML can be classified into supervised and unsupervised learning, although there are other categories or approaches such as reinforcement and transfer learning.

In *supervised learning*, data that have been labeled are fed into the ML algorithm for training. Once the algorithm has been trained, it is called a *model*. One of the greatest challenges in developing a high-quality model is being able to obtain a substantial amount of data (ie, usually thousands of instances) that have been accurately labeled. An example of supervised learning would be an algorithm trained on a dataset of chest X-rays and corresponding diagnoses [4-6]. In *unsupervised learning*, unlabeled data are fed into an algorithm, and the algorithm discovers relationships and groupings for itself. This difference is illustrated in Figure 2. An example of unsupervised learning is work in which an algorithm identified distinct clusters or groups of patients who had COVID-19 [7,8].

**Figure 2.** Illustration of (A) supervised and (B) unsupervised learning.

| Index | Data | Label |
|---|---|---|
| Data point 1 | | Circle |
| Data point 2 | | Square |
| Data point 3 | | Circle |
| Data point 4 | | Square |
| Data point 5 | | Square |
| Data point 6 | | Circle |

Figure 2(A)

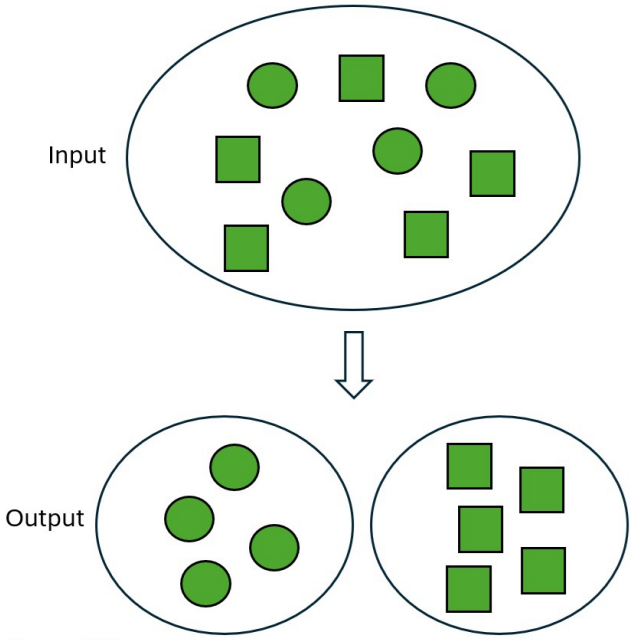



Figure 2(B)

In Figure 2A, the shapes are labeled as circles and squares by a human. The dataset is fed into the algorithm, and through these labels, the machine learns which features contribute most to classifying the shapes. In Figure 2B, the machine is fed raw data without any labels. Through exploration and differences detected in the data among its features, the model learns that there are potentially 2 different categories of objects. Of note is that the model may not inherently recognize them as a circle or a square but rather as 2 distinct categories of objects.

Other forms of learning also exist, such as *reinforcement learning*. In reinforcement learning, an algorithm experiences an environment and takes an action. On the basis of this action, the model is given a reward or a punishment as feedback. The algorithm learns which patterns result in rewards vs punishments and adjusts its behavior accordingly. Furthermore, supervised and unsupervised learning exist on a continuum, and some forms of learning are a mixture of both (ie, some instances are labeled, while others are not). A full description of these concepts is beyond the scope of this paper.

Transfer learning is another method in which an algorithm is first trained on a very large but nonspecific set of data for the desired outcome and then uses the learned patterns on a smaller but more specific set of data to refine the algorithm into a model. Transfer learning is often used where high-quality labeled datasets specific to the subject area are limited, whereas broader, nonspecific datasets are more plentiful.

## Step 4: Defining DL and Neural Networks

A subcategory of ML is DL. This branch of ML is heavily focused on neural networks. *Neural networks* were first described in the mid-20th century [9,10] and were designed to emulate neural processes in the human nervous system. While foundational work has been ongoing for decades in the field



XSL•FO
RenderX

[11-13], much early work was constrained by hardware and data availability. The emergence of powerful parallel computing, called graphics processing units; platforms to leverage graphics processing units; and the availability of large datasets have helped overcome these barriers [14].

Artificial nodes, called *neurons*, are connected in layers to form a network. Data are put into the input layer of the network; the network processes the data through one to many hidden layers and then provides results in the output layer.

DL specifically refers to ML done on neural networks with many layers (thus the term "deep"). There is no exact number of layers that is generally agreed upon as a threshold. However, common examples such as ResNet (Microsoft) and the architecture underlying ChatGPT (OpenAI) can contain a hundred layers and billions of parameters [15,16].

## Step 5a: Walk-Through on How ML Algorithms Work

Next, this tutorial will lay out the process by which ML algorithms work, describe the process with an example, and then define other key terms. To learn how to properly predict, classify, or rank instances in supervised learning, an algorithm analyzes the data to determine which features contribute the most to the data labels. This learning process is called *training*. During training, the machine learns to adjust internal parameters, called *weights*, to produce the desired outputs. These weights correspond to individual neurons. This occurs through minimizing loss functions, in which the gap between the observed and expected prediction is minimized.

For example, consider a model designed to estimate the risk of candidiasis in hospitalized patients. The model may learn that features such as intensive care admission, fever, abdominal symptoms, or a normal white blood cell count are associated with either a higher or lower risk of candidiasis. Features will often be mapped in a complex fashion across a series of neurons and layers. Modification of the weights applied to each of these features influences the resulting model's output through complex interactions across many layers of analysis, allowing the model to learn patterns that relate input features to the target outcome. In simpler forms of ML, it is sometimes possible to determine which features most greatly contribute to an outcome; however, in DL, this is often difficult because of the number of nodes and layers involved. Therefore, in DL, the change in importance may not be directly interpretable.

Datasets are often large and contain many features, some of which logically have nothing to do with the outcome being examined. If the data used to train an algorithm contain these completely unrelated variables, the algorithm may make nonsensical associations, which can later result in spuriously wrong results. For example, it may associate the color of a hospital gown and/or patient's sandwich preference with overall length of stay. However, if there are different colors of hospital gowns or menu choices for patients in intensive care units vs regular floors that are not captured in the data, then the model may miss the confounding variable completely. These errors in model development are hard to detect in all ML and even harder in DL.

## Step 5b: Walk-Through of Basic ML Mathematics Using an Example—Loss Function

To illustrate the mathematics of ML training, we can use an extremely simplified example with a continuous variable in supervised learning. It is noted that most modern DL requires massive sets of data. While specifics may change in other forms of ML, this example will illustrate general concepts. In this example, the dataset includes the trough levels of a nephrotoxic drug and the resultant measured (true) estimated glomerular filtration rate (eGFR) in patients. The measured eGFR is the labeled data in this supervised ML. An appropriate ML algorithm is selected that attempts to predict the eGFR based on drug levels (Table 1).

**Table .** Sample data for a model predicting nephrotoxin toxicity.

| Nephrotoxin level (mg/L) | True eGFR[a] (y; mL/min/1.73 $m^2$) | Machine learning model's predicted eGFR ($\hat{y}$; mL/min/1.73 $m^2$) |
|---|---|---|
| 10 | 100 | 95 |
| 13 | 95 | 92 |
| 15 | 100 | 90 |
| 20 | 90 | 85 |
| 25 | 80 | 80 |
| 30 | 65 | 75 |
| 40 | 55 | 65 |

[a]eGFR: estimated glomerular filtration rate.

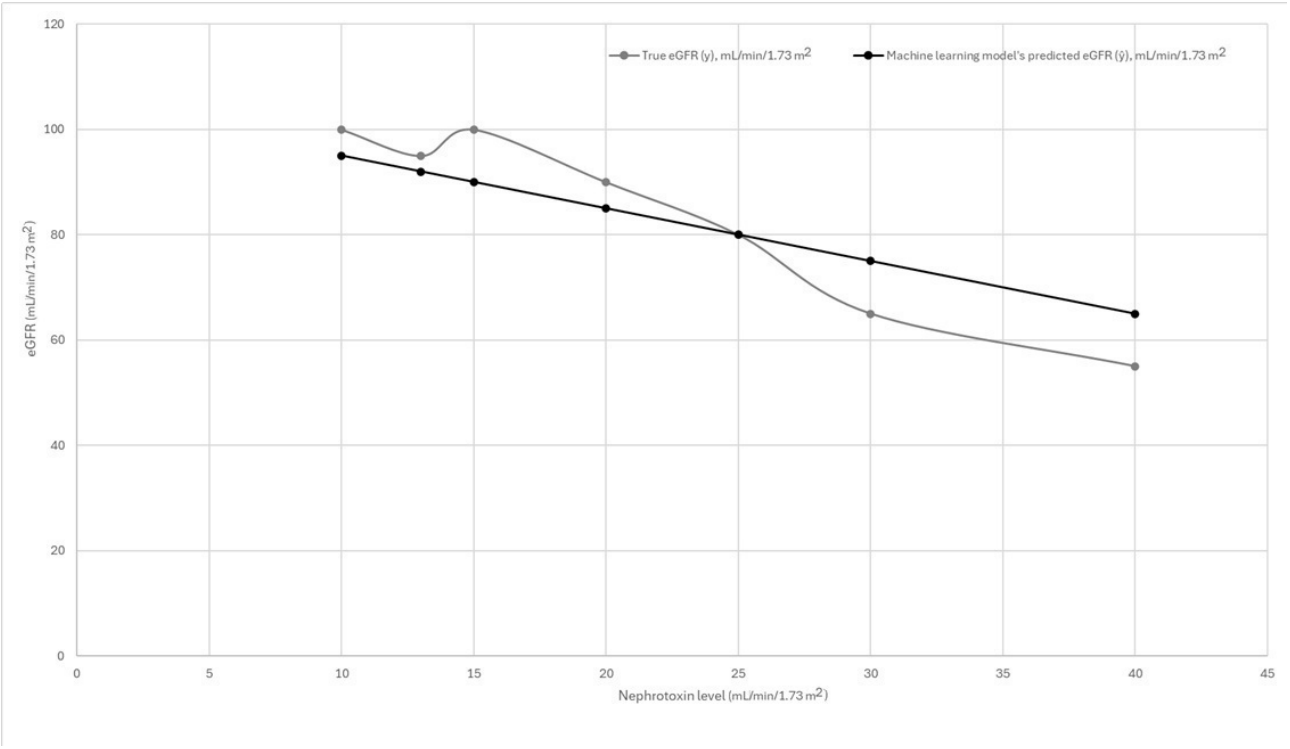
The model produces a prediction ($\hat{y}$) based on the nephrotoxin level (x). The algorithm compares the predicted eGFR ($\hat{y}$) to the real measured value of eGFR (y). Graphically, the predicted values of eGFR (the orange line in Figure 3) are plotted against the true values (the blue line). To determine the pattern between the nephrotoxin level (x) and the measured value of eGFR (y), the algorithm calculates the loss function. Differences might appear small in most regions but noticeable in some. To capture



XSL•FO
RenderX

a model's performance across all data points, we use this loss function, a mathematical function that quantifies the total error between predicted and observed values.

**Figure 3.** Sample data for a model predicting nephrotoxin level. eGFR: estimated glomerular filtration rate.




By convention, updates are made to the loss function after each batch in training. The model then adjusts the weights and tries again to determine, across all the instances in the dataset, which set of weights will produce the lowest loss function on average for the training data. With high-quality data that represent the full spectrum of possible cases that the model could be given, this helps ensure that the model is trained to perform well on average and does not result in spuriously wrong predictions. The specifics of how the algorithm calculates loss functions vary between algorithm types, with different functions being optimized for different tasks. For instance, the mean squared error, a common loss function, squares the difference between $\hat{y}$ and y. By squaring the difference, it ensures negative loss values do not cancel positive values when summed, and it helps penalize severely wrong predictions.

In a model that performs well when it is fed new data (ie, generalizes well), $\hat{y}$ and y will be similar for any new patient's nephrotoxin level fed into it. Models that do not generalize well may make large, nonsensical errors in prediction for a small subset of patients with a few specific differences in data points. This is why training a model on high-quality data that accurately represent the types of data it may encounter after deployment is critical.

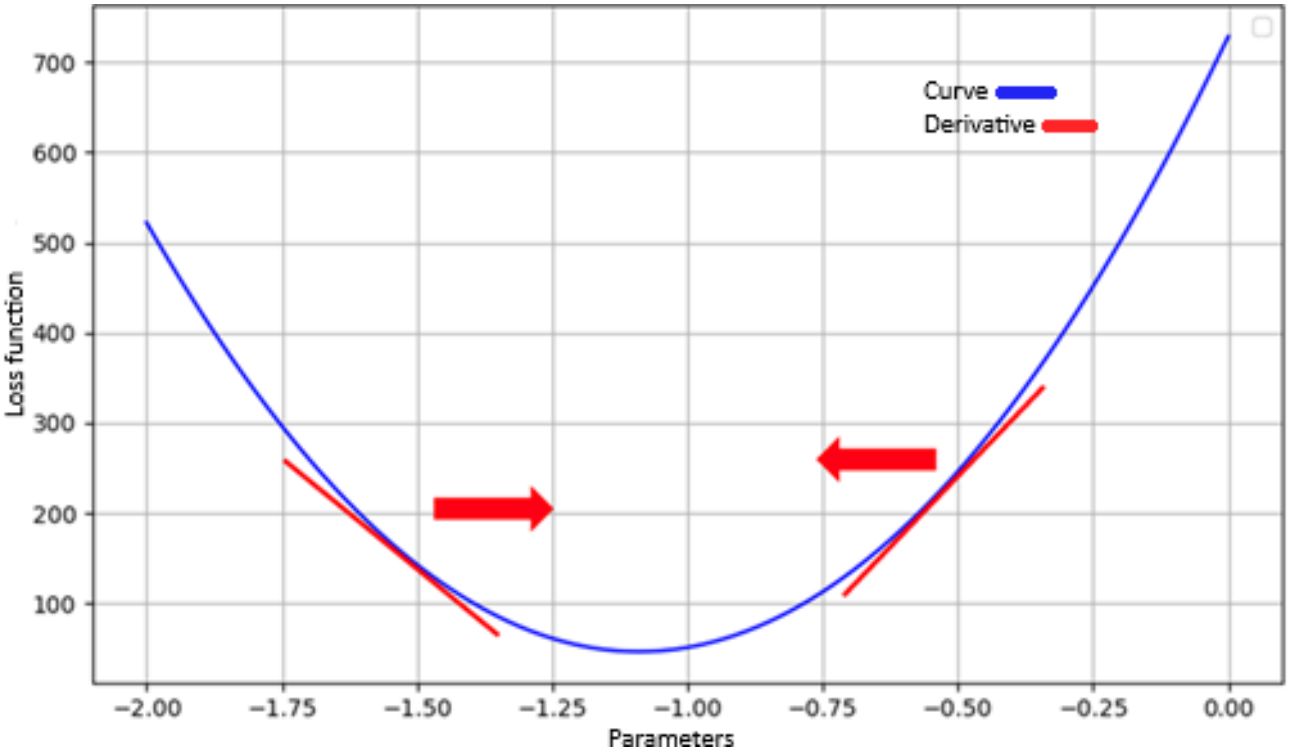
## Step 6: Understanding the Concepts of Backpropagation and Gradient Descent

Another set of key concepts to understand is backpropagation and gradient descent. In DL, a model modifies the weights of specific neurons within a neural network to create its predictions. When data are input, they forward pass through the network. The initial run creates random weights, and the prediction ability of the initial model is likely poor. However, after this forward pass, the model evaluates its resultant loss function and attempts to minimize the loss through a process called gradient descent. In gradient descent, we assign a learning rate, which increments our point along the curve of the loss function (Figure 4). If the model discovers that the loss is increasing, it will move backward to decrease the loss. Conversely, if it discovers that the loss is decreasing, it will continue moving in that direction. In doing so, it eventually seeks out the local minima, thus minimizing the loss function and improving the model's predictive capacity.



XSL•FO
RenderX

**Figure 4.** Graphical illustrations of concepts.




Mathematically, gradient descent can be expressed asw=w−η(dLdw). Remember that the derivative (dLdw) of the loss function gives the gradient at any given point. This allows the model to know if the loss is increasing or decreasing. The learning rate (η) tells the model how much to move in each direction. Recall that the loss function quantifies the difference between the predicted and observed values, with the goal of identifying the lowest point and, thus, minimizing this difference. These concepts are illustrated again visually in Figure 4. The mean squared error of our example model's simple line (Figure 3) is a parabola, creating an easy visualization of gradient descent for illustration purposes. When η is set at too large a value, it is possible that the model will jump very far back, such that it misses the local minima. Conversely, too small a value of η may result in such minimal movement that the minima is never reached.

During the training process, to specifically update the gradient, the model will undergo a process called backpropagation. The trainer algorithm will go through the neural network and alter the weights of neurons with the intention of decreasing loss. This process ultimately results in better prediction ability of the ML model.

## *Module 1: A Brief Explanation of AI Explainability and Safety*

As AI becomes significantly more powerful and integrated into society, various risks and errors have been observed. A brief discussion of safety is outlined here, and there are many other publications available that dive into each of these in detail.

How an ML model determines its output (ie, resultant prediction or result given to a user) from input is often unclear. This refers to the "black-box" nature of all ML but particularly of DL and neural networks. A field called explainable AI has emerged and attempts to not only better understand how models make their predictions but also add components to the ML model that require the model to explain how it arrived at its result (ie, which features had the most impact) [17].

For all AI, especially in medicine, it is imperative that model performance be checked for variations in outcomes that may indicate that human bias in the data has been promulgated or exacerbated by the model. AI algorithms are exquisitely sensitive pattern detection tools. As such, they can detect slight variations in data that resulted from human prejudice and bias. Worse, they can promulgate such bias into the model. Therefore, it is imperative that models be checked for differences in results that can be attributed to race, gender, socioeconomic status, religion, etc, as the presence of these elements in a model will cause the model to produce inaccurate results in certain groups of patients. Concrete examples include models having difficulty diagnosing dermatological conditions in those with darker skin tones [18], and a software program that accidentally referred White patients over African American patients to receive special care [19]. In the latter example, while the goal was to ensure patients received the care required, the algorithm was designed to predict whose care would cost more money, and it was found that less money was spent on African American patients despite having the same level of need [19]. Furthermore, a study found that an AI system could learn to predict ethnicity from radiographs alone [6]. There are now tools that assist in the detection of these elements, and models that include these elements need to be retrained on optimized data or otherwise mitigated to ensure the best care for all patients. Finally, risks associated with the infrastructure of AI also exist. As many models use a cloud-based model for computation and some public commercial large language models (LLMs) use input



XSL•FO
RenderX

data to retrain models, clinicians need to be aware of where sensitive data are being sent and stored.

Now that the use of AI is rapidly becoming more pervasive, it is reasonable to think that the degree to which AI is set to perform autonomously (ie, without a human in the middle) will increase. Some publications have discussed safety measures that are required to mitigate risks as AI becomes more autonomous and integrated into systems. These risks range from tangible risks today to theoretical risks with more powerful models. Experts have classified these risks into 4 categories: misuse, misalignment, mistakes, and structural risks. Misuse occurs when users intentionally instruct an AI tool to behave in harmful ways (ie, the user is an adversary). Misalignment occurs when AI systems knowingly act against human intent (ie, the AI is an adversary). This includes intentional deception by AI. Mistakes occur when AI systems produce incorrect outputs without intentional wrongdoing, often because real-world data are complex and influenced by many contributing factors. Finally, structural risks may emerge whereby pervasive AI systems integrated into society cause harm through the actions of multiple independent agents in a multifactorial, multiagent fashion [20].

A substantial body of ongoing research in AI focuses on ensuring and improving AI safety. For instance, studies focus on the effects of adversarial attacks on models to understand safety. Developers also attempt to build safeguards into models and use red teaming, where security professionals attempt to simulate attacks to determine robustness [4,18]. Equally important are efforts to create nuanced and well-informed regulations and guidelines for development [18,21-26].

## Module 2: Contemporary Clinical Applications of AI in Medicine

Health care is one of the most promising industries for AI. While disruptive-level work is underway, such as the ability to understand protein folding [27], AI has numerous applications currently being used routinely in health care.

Documentation is recognized to often be excessive and contributory to physician burnout, with an American Medical Informatics Association survey finding that 73.26% of health care professionals believed the time spent was inappropriate, 77.42% reported after-hours work related to documentation, and 74.83% believed that documentation impedes patient care [28]. Canadian data show similar findings, with physicians spending excessive amounts of time on administrative tasks that result in burnout [29,30].

AI scribes are tools that can ambiently listen to patient interactions and automatically generate notes for physicians, reducing the administrative burden for physicians. Numerous companies have created offerings; however, in general, scribes use LLMs, models based on the transformer architecture (which uses an attention mechanism to process preceding information and learn relationships among them) [31,32]. Within the context of scribes, the LLM uses this mechanism of self-attention to help generate logical text. Thus, the same limitations of transformers and LLMs carry forward onto AI scribes. For example, in many LLMs, hallucinations are a concern, which are theorized to arise because of algorithms being rewarded for correct responses, thereby making guessing a more advantageous response than acknowledging uncertainty. Health care providers must be aware of these limitations of AI scribes and how they may potentially arise.

LLMs have also been used by vendors such as Epic to automatically extract information out of existing notes, such as creating discharge summaries, reading radiology reports, providing summaries, preparing tasks based on the note being created, and providing insights [33]. They have also been used as chatbots to act in clerical roles and as search engines for medical knowledge [34,35]. While LLMs such as ChatGPT (OpenAI), Claude (Anthropic), and Gemini (Google) are commonly used by the general public, OpenEvidence (OpenEvidence LLC) is trained specifically on medical literature, reducing errors and hallucinations [36]. Models intended for use in research and science also exist, such as Perplexity (Perplexity AI) and Elicit (Elicit Research) [36,37]. Open source, medically-focused models without interfaces also exist, such as MedGemma (Google) [38].

Under the broader definition discussed in this paper, AI has been used in diagnostics for decades using simple rule-based algorithms. However, more recently, ML-DL–based approaches have begun to gain traction due to improved performance. Areas that have shown promise for ML-DL performance are radiology and pathology [39,40]. In radiology, a DL software such as CINA-iPE (Avicenna.AI) has shown promise in detecting pulmonary embolisms [41]. In pathology, ML has improved rapid patient diagnostics based on DNA methylation markers [42]. Sepsis and cardiac arrest prediction have been an ongoing area of work [43,44]. While results from this work are promising, due in part to interpretability, accuracy, and the potential for uncovering clinically irrelevant abnormalities (“incidentalomas”), it is unclear how clinicians should react to findings [45]. The volume of AI tool uptake is increasing, and AI will inevitably impact the workflow of clinicians in the near future. AI could also support hospital infrastructure [46].

Despite its potential benefit, AI integration into health care is still an evolving landscape. Regulation and guidance remain an important area of evolving work for health care AI, with numerous national and international bodies producing frameworks and guidelines for AI use [47-49]. Important questions that remain under debate include the responsibility for AI errors, with many institutions holding physicians ultimately accountable for clinical decisions, how to mitigate bias in health care AI propagated by inherent bias in training data, and the importance of privacy [6,18,19,50]. Furthermore, many studies have indicated that implementation into workflows can also be challenging, due to a lack of awareness and engagement by both patients and health care professionals, as well as logistical implementation challenges inherent to any health technology [51].

## Future Directions

AI promises to be one of the most critical revolutions of human society, arguably on par with the industrial or even agricultural



XSL•FO
RenderX

revolutions. AI will impact every area of human society, including health care.

While current implementations such as LLMs, predictive models such as convolutional neural networks, and the tools they have created (eg, AI scribes and ChatGPT) are influential, most of the society-changing work is being dedicated to creating artificial general intelligence and artificial superintelligence. While the exact definition of these terms, and even their possibility, is a matter of contention [52], they generally refer to AI systems that are either as intelligent as or more intelligent than human beings across a wide array of domains and tasks. In achieving this goal, AI will have implications on the role of humans in a post–artificial general intelligence society. While there are critical concerns about human displacement, there is also a potential for creating abundance, reducing scarcity, and an ability to supercharge scientific discovery [53].

Due to its importance, the authors believe it is important that clinicians receive structured educational content on the topic. Leaders could consider integrating formal foundational AI teachings into medical school curricula and, then in later years, providing a chance to discuss its implications and applications in health care. These sessions could also be incorporated into postgraduate education and into continuing medical education sessions offered by workplaces.

## Acknowledgments

The authors declare the use of generative artificial intelligence (GAI) in the research and writing process. According to the Generative Artificial Intelligence Delegation Taxonomy (2025), the following tasks were delegated to GAI tools under full human supervision: proofreading; editing; adapting and adjusting emotional tone; and reformatting sentence structure in specific, limited portions of text. On occasion, ChatGPT was used as a search engine to find links to relevant references, with authors further reviewing and ensuring the accuracy of these references. The GAI tool used was ChatGPT (version 5.1; OpenAI). Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes.

## Funding

No external financial support or grants were received from any public, commercial, not-for-profit, or other entity for any part of this work.

## Conflicts of Interest

None declared.

## References

1. What is artificial intelligence (AI)? IBM. URL: https://www.ibm.com/think/topics/artificial-intelligence [accessed 2025-06-28]
2. Shortliffe EH. A rule-based computer program for advising physicians regarding antimicrobial therapy selection. Presented at: ACM '74: Proceedings of the 1974 Annual ACM Conference; Jan 1, 1974. [doi: 10.1145/1408800.1408906]
3. Campbell M. Knowledge discovery in deep blue. Commun ACM 1999 Nov;42(11):65-67. [doi: 10.1145/319382.319396]
4. Lee SB. Development of a chest X-ray machine learning convolutional neural network model on a budget and using artificial intelligence explainability techniques to analyze patterns of machine learning inference. JAMIA Open 2024 Jul;7(2):ooae035. [doi: 10.1093/jamiaopen/ooae035] [Medline: 38699648]
5. Lee SB. Gradual poisoning of a chest x-ray convolutional neural network with an adversarial attack and AI explainability methods. Sci Rep 2025 Jul 1;15(1):21779. [doi: 10.1038/s41598-025-02294-3] [Medline: 40593872]
6. Gichoya JW, Banerjee I, Bhimireddy AR, et al. AI recognition of patient race in medical imaging: a modelling study. Lancet Digit Health 2022 Jun;4(6):e406-e414. [doi: 10.1016/S2589-7500(22)00063-2] [Medline: 35568690]
7. Benito-León J, Del Castillo MD, Estirado A, Ghosh R, Dubey S, Serrano JI. Using unsupervised machine learning to identify age- and sex-independent severity subgroups among patients with COVID-19: observational longitudinal study. J Med Internet Res 2021 May 27;23(5):e25988. [doi: 10.2196/25988] [Medline: 33872186]
8. Nalinthasnai N, Thammasudjarit R, Tassaneyasin T, et al. Unsupervised machine learning clustering approach for hospitalized COVID-19 pneumonia patients. BMC Pulm Med 2025 Feb 8;25(1):70. [doi: 10.1186/s12890-025-03536-w] [Medline: 39923003]
9. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 1958 Nov;65(6):386-408. [doi: 10.1037/h0042519] [Medline: 13602029]
10. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 1943 Dec;5(4):115-133. [doi: 10.1007/BF02478259]
11. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986 Oct;323(6088):533-536. [doi: 10.1038/323533a0]
12. Lippmann R. An introduction to computing with neural nets. IEEE ASSP Mag 1987;4(2):4-22. [doi: 10.1109/MASSP.1987.1165576]



XSL•FO
RenderX

13. Ackley DH, Hinton GE, Sejnowski TJ. A learning algorithm for Boltzmann machines. Cogn Sci 1985 Mar;9(1):147-169. [doi: 10.1016/S0364-0213(85)80012-4]
14. Li M, Bi Z, Wang T, et al. Deep learning and machine learning with GPGPU and CUDA: unlocking the power of parallel computing. arXiv. Preprint posted online on Oct 8, 2024. [doi: 10.48550/arXiv.2410.05686]
15. Liu BD, Meng J, Xie WY, Shao S, Li Y, Wang Y. Weighted spatial pyramid matching collaborative representation for remote-sensing-image scene classification. Remote Sens (Basel) 2019;11(5):518. [doi: 10.3390/rs11050518]
16. Alarcon N. OpenAI presents GPT-3, a 175 billion parameters language model. NVIDIA Developer. 2020 Jul 7. URL: https://developer.nvidia.com/blog/openai-presents-gpt-3-a-175-billion-parameters-language-model/ [accessed 2026-01-17]
17. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Presented at: Proceedings of the 31st International Conference on Neural Information Processing Systems; Dec 4-9, 2017. [doi: 10.5555/3295222.3295230]
18. Dowie T. Exploring the diagnostic capability of artificial intelligence in dermatology for darker skin tones: a narrative review. Cureus 2025 Oct;17(10):e94909. [doi: 10.7759/cureus.94909] [Medline: 41262830]
19. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
20. An approach to technical AGI safety and security. Medium. 2025. URL: https://deepmindsafetyresearch.medium.com/an-approach-to-technical-agi-safety-and-security-25928819fbc6 [accessed 2025-1-17]
21. Norgeot B, Quer G, Beaulieu-Jones BK, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. Nat Med 2020 Sep;26(9):1320-1324. [doi: 10.1038/s41591-020-1041-y] [Medline: 32908275]
22. Predetermined change control plans for machine learning-enabled medical devices: guiding principles. U.S. Food & Drug Administration. 2025. URL: https://www.fda.gov/medical-devices/software-medical-device-samd/predetermined-change-control-plans-machine-learning-enabled-medical-devices-guiding-principles [accessed 2026-02-25]
23. Transparency for machine learning-enabled medical devices: guiding principles. U.S. Food & Drug Administration. 2024 Jun 13. URL: https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles [accessed 2026-02-25]
24. Good machine learning practice for medical device development: guiding principles. U.S. Food & Drug Administration. 2025. URL: https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles [accessed 2026-02-25]
25. Engineering fundamentals checklist. Microsoft Open Source. URL: https://microsoft.github.io/code-with-engineering-playbook/engineering-fundamentals-checklist/ [accessed 2026-01-17]
26. Koh RGL, Khan MA, Rashidiani S, et al. Check it before you wreck it: a guide to STAR-ML for screening machine learning reporting in research. IEEE Access 2023;11:101567-101579. [doi: 10.1109/ACCESS.2023.3316019]
27. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021 Aug;596(7873):583-589. [doi: 10.1038/s41586-021-03819-2] [Medline: 34265844]
28. Sloss EA, Owoyemi A, Mishra AK, et al. Development of the TrendBurden survey: assessing perceived documentation burden among health professionals in the United States. Appl Clin Inform 2025 May;16(3):662-675. [doi: 10.1055/a-2562-0910] [Medline: 40669861]
29. Physician administrative burden survey –final report. : Doctors Nova Scotia; 2020 Sep URL: https://doctorsns.com/sites/default/files/2020-11/admin-burden-survey-results.pdf [accessed 2026-02-25]
30. Joint Task Force to reduce administrative burdens on physicians. Doctors Manitoba. 2023 May 30. URL: https://assets.doctorsmanitoba.ca/documents/Admin-Burden-Progress-Report-May-30.pdf [accessed 2026-02-25]
31. Acallar LJ. AI medical scribes: everything you need to know. Heidi. 2026. URL: https://www.heidihealth.com/blog/ai-medical-scribe?utm_source=chatgpt.com [accessed 2026-02-25]
32. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Presented at: 31st Conference on Neural Information Processing Systems (NIPS 2017); Dec 4-7, 2017 URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf [accessed 2026-02-25]
33. AI for clinicians. Epic. URL: https://www.epic.com/software/ai-clinicians/ [accessed 2026-02-25]
34. AI receptionist answers the call for busy medical clinics. Hamilton Health Sciences. URL: www.hamiltonhealthsciences.ca/share/patients-test-ai-receptionist [accessed 2026-03-20]
35. Open Evidence. URL: www.openevidence.com [accessed 2026-03-20]
36. AI for scientific research. Elicit. URL: https://elicit.com [accessed 2026-02-25]
37. Perplexity. URL: https://www.perplexity.ai [accessed 2026-02-25]
38. Google DeepMind. MedGemma. URL: https://deepmind.google/models/gemma/medgemma [accessed 2026-02-25]
39. Lawrence R, Dodsworth E, Massou E, et al. Artificial intelligence for diagnostics in radiology practice: a rapid systematic scoping review. EClinicalMedicine 2025 May 12;83:103228. [doi: 10.1016/j.eclinm.2025.103228] [Medline: 40474995]
40. Cazzato G, Rongioletti F. Artificial intelligence in dermatopathology: updates, strengths, and challenges. Clin Dermatol 2024;42(5):437-442. [doi: 10.1016/j.clindermatol.2024.06.010] [Medline: 38909860]
41. Farzaneh H, Junn J, Chaibi Y, et al. Deep learning-based algorithm for automatic detection of incidental pulmonary embolism on contrast-enhanced CT: a multicenter multivendor study. Radiol Adv 2025 Jun 23;2(4):umaf021. [doi: 10.1093/radadv/umaf021] [Medline: 41058960]




XSL•FO
RenderX

42. Aref-Eshghi E, Abadi AB, Farhadieh ME, et al. DNA methylation and machine learning: challenges and perspective toward enhanced clinical diagnostics. Clin Epigenetics 2025 Oct 10;17(1):170. [doi: 10.1186/s13148-025-01967-0] [Medline: 41074112]
43. Chang WS, Hsiao KY, Lin LY, Chen M, Shia BC, Lin CY. Machine learning models for predicting in-hospital cardiac arrest: a comparative analysis with logistic regression. Int J Gen Med 2025;18:6341-6352. [doi: 10.2147/IJGM.S569559] [Medline: 41141891]
44. Drysch M, Reinkemeier F, Puscz F, et al. Streamlined machine learning model for early sepsis risk prediction in burn patients. NPJ Digit Med 2025 Oct 21;8(1):621. [doi: 10.1038/s41746-025-02078-z] [Medline: 41120704]
45. Moss L, Corsar D, Shaw M, Piper I, Hawthorne C. Demystifying the black box: the importance of interpretability of predictive models in neurocritical care. Neurocrit Care 2022 Aug;37(Suppl 2):185-191. [doi: 10.1007/s12028-022-01504-4] [Medline: 35523917]
46. Kumar AK, Ali Y, Kumar RR, Assaf MH, Ilyas S. Artificial intelligent and internet of things framework for sustainable hazardous waste management in hospitals. Waste Manag 2025 Jul 15;203:114816. [doi: 10.1016/j.wasman.2025.114816] [Medline: 40311410]
47. Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240029200 [accessed 2026-02-25]
48. The medico-legal lens on AI use by Canadian physicians. Canadian Medical Protective Association. 2024 Sep. URL: https://www.cmpa-acpm.ca/en/research-policy/public-policy/the-medico-legal-lens-on-ai-use-by-canadian-physicians [accessed 2026-02-25]
49. Solomonides AE, Koski E, Atabaki SM, et al. Defining AMIA's artificial intelligence principles. J Am Med Inform Assoc 2022 Mar 15;29(4):585-591. [doi: 10.1093/jamia/ocac006] [Medline: 32722749]
50. Artificial intelligence (AI) in medical practice. College of Physicians and Surgeons of Saskatchewan. URL: https://www.cps.sk.ca/imis/ContentBuddyDownload.aspx?DocumentVersionKey=879928d2-6caa-4894-a866-86fbea85ce57 [accessed 2026-01-19]
51. Livieri G, Mangina E, Protopapadakis ED, Panayiotou AG. The gaps and challenges in digital health technology use as perceived by patients: a scoping review and narrative meta-synthesis. Front Digit Health 2025 Mar 27;7:1474956. [doi: 10.3389/fdgth.2025.1474956] [Medline: 40212901]
52. Morris MR, Sohl-Dickstein J, Fiedel N, et al. Levels of AGI for operationalizing progress on the path to AGI. : arXiv Preprint posted online on Nov 4, 2023. [doi: 10.48550/arXiv.2311.02462]
53. Xu Y, Liu X, Cao X, et al. Artificial intelligence: a powerful paradigm for scientific research. Innovation (Camb) 2021 Oct 28;2(4):100179. [doi: 10.1016/j.xinn.2021.100179] [Medline: 34877560]


## Abbreviations

**AI:** artificial intelligence
**DL:** deep learning
**eGFR:** estimated glomerular filtration rate
**LLM:** large language model
**ML:** machine learning


*Edited by A Schwartz, M Balcarras; submitted 04.Oct.2025; peer-reviewed by F Fukuzawa, RR Kumar; revised version received 05.Feb.2026; accepted 06.Feb.2026; published 30.Mar.2026.*

*Please cite as:*
*Lee SB, Carter AB, Haider MH, Ko SB*
*An Introduction to AI for Clinicians: Tutorial*
*Interact J Med Res 2026;15:e85266*
*URL: https://www.i-jmr.org/2026/1/e85266*
*doi:10.2196/85266*


© Stephen B Lee, Alexis B Carter, Muhammad Hamis Haider, Seok-Bum Ko. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 30.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Enhancing Transparency, Auditability, and Reproducibility of Deduplication in Systematic Reviews: Tutorial for the Rayyan Method and Systematic Auto Resolver

Mimi M Kim, PhD; Amin M Shanaa, BSc; Hossam Hammady, BSc, MEng; Mohammad Aboelnour, BSc; Robert Ayan, BSBA, MBA

Rayyan Systems Inc., 1 Broadway, 14th Floor, Cambridge, MA, United States

**Corresponding Author:**
Mimi M Kim, PhD
Rayyan Systems Inc., 1 Broadway, 14th Floor, Cambridge, MA, United States

## Abstract

Deduplication across search results is one of the earliest and most critical steps in the systematic review methodological process; yet, existing solutions often lack the transparency, auditability, and reproducibility required by rigorous systematic review standards. Many automated deduplication tools introduce bias through opaque, nonconfigurable algorithmic decisions, while also potentially removing relevant references through false positive identification. We provide a tutorial on the Rayyan Method, including the Systematic Auto Resolver feature for the deduplication process. This method is defined by an enhanced deduplication approach that combines high-sensitivity duplicate detection with user-controlled resolution criteria. Systematic Auto Resolver enables research teams to define, apply, and document their own deduplication standards rather than relying on predetermined automated decisions. Users apply deduplication criteria, iteratively reviewing results after each pass, maintaining complete control over methodological decisions. The Rayyan Method addresses critical limitations in current deduplication approaches by supporting methodological rigor through user-controlled resolution while enhancing efficiency, transparency, and reproducibility. By empowering research teams to define their own deduplication criteria, this approach supports and aligns with the prescribed methodologically rigorous systematic review process. The method provides a citable framework for researchers to comprehensively document their deduplication methodology.

***(Interact J Med Res 2026;15:e94317)*** doi:10.2196/94317

**KEYWORDS**

systematic review methodology; deduplication; software demonstration; transparency; evidence synthesis

## Introduction

The escalating demand for evidence-based decision-making across diverse sectors, notably health care and policy, has amplified the reliance on methodologically rigorous systematic reviews for robust evidence synthesis [1]. The systematic review process begins with a carefully constructed literature search strategy, adhering to predefined, transparent, and registered criteria [2]. Central to maintaining the rigor of systematic reviews is the effective deduplication of search results, a crucial step to minimize reviewer workload, mitigate bias, and ensure the accuracy of findings [3].

While many software platforms offer a deduplication feature, their functionalities can lack the full integration, auditability, transparency, and reproducibility that are essential for regulatory applications [4,5]. With the increasing automation of research workflows, concerns have grown commensurately about the critical need to maintain controlled, transparent processes in regulatory contexts, particularly with the application of artificial intelligence. However, many researchers have adopted deduplication solutions that may inadvertently remove or cause the removal of relevant evidence in pursuit of identifying the most complete set of duplicates. This is inherently contradictory and perilous to the methodology.

## The Multidimensional Challenge of Defining Duplicate References

The primary challenge of this stage of the review stems from the absence of a consistent definition of what constitutes a duplicate in an evidence synthesis context. For example, duplicates have been defined as “the same article published in the same place, while the same article published in a different place is not a duplicate” [6]. In contrast, duplicate records have been described as “being the same bibliographic record (irrespective of how the citation details were reported; eg, variations in page numbers, author details, accents used, or abridged titles)” [7,8]. Additional factors that may result in duplicate records include intradatabase and interdatabase duplication due to differences in publication year, volume, or page numbers resulting from early release or online-first



XSL•FO
RenderX

publication [9]. These examples of varied definitions highlight the need for deduplication solutions to address the primary challenge by accommodating the varying criteria used to define duplicate references.

A critical dichotomy has emerged with the increasing number of available solutions: (1) the failure to detect duplicates with single-character deviations or greater (where missing, incomplete, and erroneous data cause duplicates to be missed entirely) and (2) the failure to avoid false positives (where highly similar references are erroneously resolved as duplicates). The result can be significantly impactful when the removal of relevant evidence identified as false positives compromises the integrity of systematic reviews. This can raise concerns about biased conclusions stemming from omitted evidence, where any research produced using such a tool would be suspected of the same. Consequently, there is a critical need for testing, auditing, and validation of deduplication methods and solutions. When solutions offer a single method of deduplication, the method, concerningly, minimizes human expertise in the method's application [7,10].

Rayyan's approach is designed to address this duplication complexity by fulfilling 2 research paradigms. First, convenience, where researchers rely on automated tools for deduplication while recognizing the dependency on the applied method or algorithm and the data assessed. Second, integrity, where researchers follow transparent, auditable, and reproducible procedures to ensure methodological rigor when removing inexact duplicates.

## History and Evolution

Rayyan was originally conceived with the objective of supporting the most time-consuming aspects of systematic literature reviews by enhancing the process of deduplication and screening of titles and abstracts for researchers with varying experience and expertise across a broad spectrum of review types (eg, literature reviews, scoping reviews, and systematic reviews). Initially, Rayyan offered automatic duplicate detection; however, after realizing that automatic detection violated the platform's mission to protect full user control, Rayyan updated the design to execute a semiautomated process with a foundational commitment to a level of usability governed by speed, accuracy, and simplicity while maintaining full user control.

We, as the Rayyan team, aimed to design a comprehensive systematic review application enabling human-computer research collaboration where research teams have a solution that conforms to their unique requirements, methods, standards, and even individual preferences. Rayyan's user-centric design philosophy is driven by our commitment to methodological rigor without imposing a single *one-size-fits-all* approach to systematic reviews, including the important and sensitive challenge of deduplication. Further details on the history of Rayyan's inception, architecture, usability, accuracy against manual methods, and the added value of Rayyan's other features are published elsewhere [10].

## Objective

This tutorial specifically describes the deduplication process using the unique capabilities of Systematic Auto Resolver and its advantages. Together, the capabilities of Rayyan combined with the user-defined deduplication criteria constitute a deduplication method, henceforth referred to as the Rayyan Method.

## The Rayyan Method

### Background

The Rayyan Method is guided by the principle of implementing human control and oversight rather than a human-in-the-loop paradigm. In human-in-the-loop systems, humans are often informed of the results after automated actions occur [11]. The Rayyan Method instead starts with human initiation and guidance of deduplication decisions, preventing the automated removal of potentially relevant evidence. This distinction ensures that humans initiate the action and guide it with human expertise.

The pioneering work of methodologist Bramer established the concept of a manual series of deduplication criteria to achieve superior deduplication results beyond exact duplicates [4]. The Rayyan Method also meets the same objectives and extends the process by providing researchers with the capability where deduplication criteria are defined by the user, rather than predetermined by software developers. This evolution represents a shift from fixed procedural steps to adaptable, user-controlled criteria that can accommodate the diversity of research contexts and requirements.

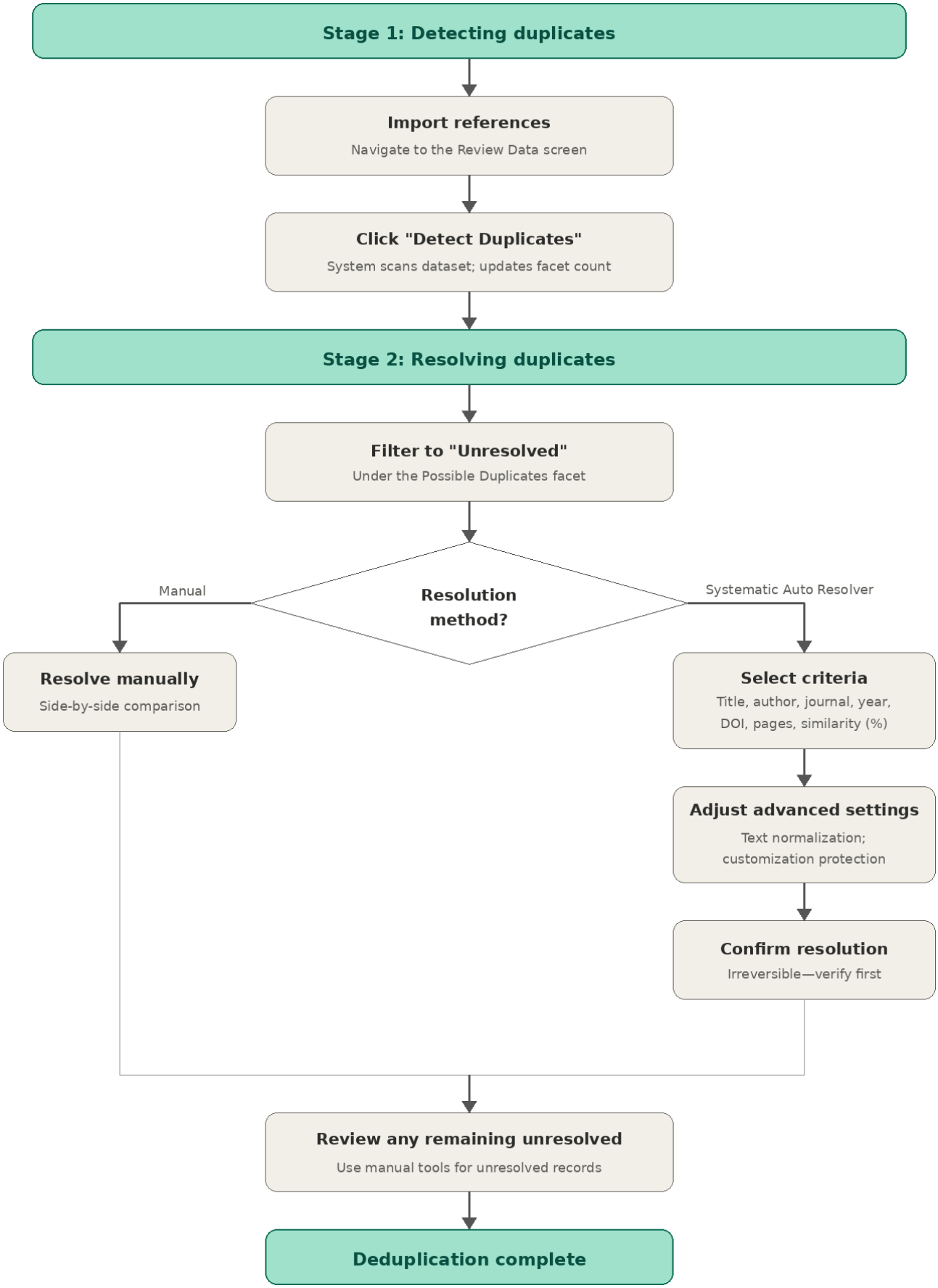
Rayyan offers an advanced duplicate detection algorithm within both its web-based and mobile app designed for the systematic review process [10]. Using greater sensitivity, Rayyan detects more possible duplicates than other solutions [7]. Using precision rules, Rayyan minimizes false positives and false negatives by allowing the review team to specify which deduplication criteria to apply to their dataset. The applied deduplication criteria can be reported for transparent, auditable, and reproducible reporting of the deduplication method. Any possible duplicates that are not removed by the application of the deduplication criteria can be manually reviewed and resolved. Manual resolution is most common for instances of multiplicity stemming from high similarity, where the research team is then able to decide whether to keep one or both references.

This is a 2-stage process that separates detection from resolution (Figure 1). This separation is critical because it allows for maximum sensitivity in detection without the risk of false positives in resolution. In stage 1, Rayyan casts a wide net to identify all possible duplicates to delegate precision to stage 2 resolution. In stage 2, research teams apply their own criteria to resolve these duplicates systematically and transparently.



XSL•FO
RenderX

**Figure 1.** Deduplication and Systematic Auto Resolver step-by-step process.






XSL•FO
RenderX

### Stage 1: Detecting Duplicates

Research teams start by importing all their search files containing the references that constitute the possible evidence base for their systematic review into the Rayyan platform. Rayyan also offers a *PDF first* import option where the platform creates the references and retrieves metadata based on the initial PDF upload. Rayyan currently supports importing 1 GB for ZIP files, 100 MB for reference files, and 10 MB for PDFs, with support for a large array of industry-standard content formats (EndNote, RefMan, BibTeX, PubMed, Web of Science, PDF, and more) and file formats (.enw, .ris, .bib, .csv, .wos, .txt, .xml, .doc, .zip, and .pdf). The simultaneous import of more than 10 search files may be repeated any number of times to add additional references to a review for processing in real time. Any limitation on the size of the search files, the number of search files that may be simultaneously imported, or the number of references that duplicate detection can handle has been specified as a safeguard to prevent misuse of the platform, given that duplicate detection is a free feature of Rayyan available to anyone. However, users may also overcome the 10-file limit by importing searches from Rayyan's "My Library" to a review.

Once all references are uploaded to Rayyan by the review team, the deduplication process is user initiated with one click of the Detect Duplicates button, returning a set of records identified as "Possible Duplicates" as a filtered index of records. The Detect Duplicates button may be found in several locations, including the Overview page, Review Data page, or the Rayyan mobile app. Initiating the duplicate detection process from any of these interfaces or via the application programming interface has the same effect. Duplicate detection and removal are user initiated, meaning deduplication is neither automatic nor required to use Rayyan to complete an end-to-end systematic review workflow.

Rayyan's detection algorithm identifies similar records [10] to generate "Possible Duplicates," defined as having sufficient similarity to other records. The returned results include a set of possible duplicate records with their individual similarity scores presented in an additional column. In independent third-party evaluations, Rayyan possessed remarkably high and consistent accuracy, sensitivity, and specificity for identifying duplicate references [7,12].

### Step 2: Resolving Duplicates

#### *Selecting Deduplication Criteria*

After executing Detect Duplicates, the end user is provided with a similarity score (percentage of similarity) for records sharing similarities. The end user is presented with the option to manually resolve or autoresolve these potential duplicates based on the criteria defined a priori. The Systematic Auto Resolver empowers the end user with control over the criteria used to define duplicates within the set of detected possible duplicates. For example, a review team might choose title, author, year of publication, and DOI to ensure accuracy. Once the universe of possible duplicates has been identified, the process of resolving duplicates manually and with the Systematic Auto Resolver is enabled.

The deduplication criteria available for selection include the following:

- Title
- Author
- Journal
- Year
- Pages
- DOI
- Publication type
- Overall similarity score (percentage of similarity based on title, journal, year, and author information)

These criteria can be applied individually or in combination. The end user specifies the deduplication criteria that are applied to resolve duplicates. This level of control allows for tailored deduplication strategies that are not possible with fully automated tools.

#### *Manual Resolution of Duplicates*

The first stage of duplicate resolution is manual resolution. By selecting the "Unresolved" filter under the "Possible Duplicates" filter on the "Review Data" screen, a list of possible duplicate records is shown. By selecting a record listed within the Unresolved filter and clicking "Resolve Duplicates," paired combinations of possible duplicates, in plurality for those instances, are presented side by side for the end user to resolve manually. This manual process intends to provide maximum control and human oversight over the verification and removal of duplicates from the systematic review. The steps for manual resolution of duplicates are as follows:

1. Visit the "Review Data" screen
2. Click "Detect Duplicates" to complete duplicate detection
3. Select "Unresolved" from the Possible Duplicates facet
4. Review and resolve duplicates manually by selecting which records to keep (left, right, or both) or marking them as "Not Duplicate"

To export the deleted records, select the "Deleted" filter in the Possible Duplicates facet. Then select "Export" and select "Filtered" to export only the deleted records for reporting.

#### *Systematic Auto Resolver Feature for Duplicate Resolution*

The second stage of duplicate resolution is to apply Systematic Auto Resolver and is summarized in the following steps:

1. Access the Systematic Auto Resolver
2. Select the deduplication criteria
3. Adjust advanced settings
4. Confirm autoresolution
5. View notifications and results

Research teams may repeat the application of the Systematic Auto Resolver with a progressive series of rules from the most conservative to the most encompassing or resort to manual resolution to deal with unresolved possible duplicates that are resistant to the application of rules. The ability to control and exercise oversight is intended to minimize the risk of undetected false positives that can occur when single-stage automated tools delete high-similarity possible duplicates as duplicates without



XSL•FO
RenderX

user authorization or visibility. It is important to note, however, that the Systematic Auto Resolver is not immune to false positives, particularly when few criteria are applied without sufficient specificity or methodological grounding. The iterative, criteria-based approach described in this tutorial is designed to reduce this risk, and the auditable record of all resolution decisions ensures that any errors remain traceable and correctable by the research team.

While concerns have heightened over the increasing automation of systematic reviews, with particular sensitivity around the behavior of artificial intelligence, many ignore the seemingly innocuous endorsement of single-stage deduplication tools that may cause the removal of relevant evidence without human control or oversight. Like other early errors in the systematic review process, such as poor protocol development, deficient search strategies, or poor search implementation, deleting relevant evidence using an automation tool jeopardizes the integrity of the systematic review process from the earliest steps. Tool adoption becomes an implicit endorsement of the deduplication method it uses. Unlike single-stage automated tools that remove duplicates without user authorization or visibility, the Rayyan Method ensures that resolution decisions, whether manual or criteria-based, are explicitly authorized, documented, and auditable by the research team. This 2-step deduplication process does not entirely eliminate the possibility of false positives but ensures that any such errors are traceable to specific user-defined criteria rather than opaque algorithmic decisions, preserving the researcher's ability to identify, review, and correct duplicate cases.

However, user control is only as rigorous as the methodological decisions informing it. Transparency in tooling does not substitute for competency in application; instead, it transfers responsibility to the reviewer. This is a meaningful tradeoff: where closed, automated tools make implicit methodological decisions on the user's behalf, the Rayyan Method requires users to make those decisions explicitly and deliberately. For researchers without deep familiarity with deduplication methodology, this degree of control can itself become a source of error if criteria are selected without an understanding of their relative reliability, limitations, or appropriate sequencing (eg, applying DOI as a sole matching criterion can systematically overlook duplicates in records with incomplete or "dirty" metadata). The considerations that follow are designed to address this directly by providing not only procedural steps but also the methodological rationale behind each decision point, including known field-specific limitations that reviewers should consider before proceeding. Readers seeking a broader evidence base for criteria selection are encouraged to consult the methodological literature on deduplication field performance [4,6,7,13].

## Additional Considerations

### *Options for Mitigating Risks of False Positives and False Negatives When Using Systematic Auto Resolver*

Optimizing the selection of criteria requires an understanding that no single field is perfectly reliable across all records or databases [4]. There is variation across all the criterion fields in how they are populated and/or formatted, which directly impacts sensitivity (detecting more true duplicates) and specificity (avoiding the removal of unique records). The evidence base includes literature on the tradeoffs between sensitivity and specificity and should be considered when configuring a criterion approach [4,7]. We recommend selecting criteria based on published validation evidence rather than ad hoc testing. The methodological literature provides empirically grounded guidance on the relative reliability of individual matching fields [4,7,8], which reviewers should consult when configuring Systematic Auto Resolver passes. A stepwise approach, beginning with high-specificity combinations and progressively broadening, has been shown to maximize duplicate detection while minimizing the risk of false-positive removal. To inform the definition of this approach, we provide the following considerations for the various criterion fields: Title; Year, Author, and Journal; DOI; Pages and Publication Type; and Overall Similarity Percentage.

*Title* is a consistently populated and stable field across databases but also has varying representations across databases. With Rayyan, text normalization, which converts text to lowercase and removes punctuation (discussed further below), should be enabled to execute title-based matching and minimize minor formatting differences between database records that can lead to true duplicates being missed.

*Year*, *Author*, and *Journal* are criteria that can serve as confirmatory criteria used in conjunction with Title rather than as standalone matching fields. Year is a supplement that can reduce false positives from a similar titled but distinct articles, thereby reducing false positives. Author fields commonly have variations across databases (eg, full name vs initials, surname order), which can reduce recall when used alone as an identifier. Journal fields are reliable when well indexed in biomedical literature and less so for gray literature, conference proceedings, or preprints. Overall, these fields are subject to variation across databases that can significantly impact duplicate detection and resolution.

*DOI* is considered a highly precise criterion but with lower recall. DOIs that are confirmed as matched can be reliably considered true duplicates; however, DOI metadata can be limited or erroneous for a proportion of records [14,15]. Consequently, we recommend DOI as a supplementary confirmatory criterion in later iterations and not as a primary or sole matching criterion.

Additionally, *Pages and Publication Type* are also less reliable as standalone criteria. An important consideration is that the formatting of page numbers frequently varies across databases (eg, first page only vs full range; online-first vs final print pagination). The same is often the case for Publication Type. Therefore, these fields are potentially most useful as supplementary criteria within multifield combinations rather than as primary matching criteria.

Finally, as reported in the Rayyan platform, the *Overall Similarity Percentage* is potentially most useful for capturing duplicates with inconsistent metadata across multiple fields. Higher thresholds of ≥95% approximate exact matching with tolerance for minor variation and carry a lower risk of false



XSL•FO
RenderX

positives. Lower thresholds increase the number of duplicates detected but require more careful manual review of flagged records. We recommend beginning at ≥95% and reviewing the results before considering a lower threshold.

### *Applying Criteria Iteratively*

Rather than applying all criteria in a single pass, we recommend a stepwise approach grounded in the evidence base, as established by Bramer et al [4], among others [15]. Evidence strongly supports that applying a multistep method that begins with high-specificity field combinations and then broadens the criteria can outperform single-pass deduplication approaches [4]. With this evidence-based guidance in mind, Rayyan's Systematic Auto Resolver is designed to support sequential passes with criteria sets. An example of a practical evidence-based methodological sequence for many review types could be as follows:

- Pass 1: Title + Year + Author. This combination captures the majority of exact and near-exact duplicates with minimal risk of false positives. It is the appropriate starting point for most datasets.
- Pass 2: Title + Journal, or Title alone with text normalization enabled. This pass targets records with author field inconsistencies or missing author metadata that may have been missed in Pass 1.
- Pass 3: Previous fields + DOI + Pages + Publication Type. A final high-confidence pass for any remaining unresolved records where additional metadata are present and verified.
- Pass 4: Overall Similarity at ≥95%. Review flagged records before confirming resolution, as this pass is most likely to surface borderline cases requiring judgment.

Rayyan does not take any action specifically not authorized by end users. This means the integrity of the results is entirely in the hands of the researcher, where the application of methodological rigor guarantees integrity. Rayyan's role in this instance is to apply the tools as instructed by the user to produce the results. It is important to note that because Systematic Auto Resolver decisions are irreversible, prioritizing specificity in early passes and reserving broader criteria for later passes, where manual review of flagged results is feasible, reduces the risk of inadvertently removing unique records.

Tool transparency does not guarantee review transparency. Rayyan's Systematic Auto Resolver provides a transparent and auditable process, but this only contributes to a reproducible review record if researchers actively document and report their deduplication decisions. In accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 reporting standards [1] and best practices for deduplication reporting [4,7,13], researchers should document the following elements in their methods section, registered protocol, or supplementary materials:

- Tool and date: The name of the deduplication software used and the date on which deduplication was performed. Software features and criteria may be updated over time, so the date of use contextualizes the methodology for future readers.
- Detection method: Whether duplicate detection was initiated via the Detect Duplicates function and any settings applied at that stage.
- Criteria per pass: For each Systematic Auto Resolver pass, the specific criteria applied (eg, Title + Year + Author; Overall Similarity ≥95%), the sequence of passes, and whether text normalization or customization protection was enabled.
- Resolution counts: The total number of references flagged as possible duplicates; the number resolved via Systematic Auto Resolver; the number resolved manually; and the number remaining unresolved after all passes.
- Criteria rationale: A brief statement of the reasoning behind criteria selection, particularly where choices were informed by dataset-specific factors or deviated from standard stepwise practice.

Research teams that have conducted deduplication partially or entirely outside Rayyan should note that Rayyan's PRISMA flow diagram generator allows *n* values to be manually overridden to ensure the completed diagram accurately reflects the full deduplication process, regardless of where it was performed. This should be documented in the methods section alongside the other deduplication reporting elements listed above. Thorough documentation of these elements ensures that the deduplication process is fully citable, replicable by other research teams, and auditable by journal reviewers and editors.

## Advanced Settings

In addition, the following advanced settings may be applied as part of the deduplication criteria.

*Select Default Search*: The end user may specify which imported dataset to preserve when duplicates are detected across different searches. This setting is applied independently for each Systematic Auto Resolver pass, meaning researchers can select a different default dataset across passes to reflect the relative quality or completeness of metadata from different sources for a given criteria set. For any duplicates that remain unresolved after Systematic Auto Resolver passes, the user retains full control at the individual record level during manual review, where they may choose whichever version is most appropriate regardless of any default dataset previously specified. This flexibility may benefit the overall systematic review workflow, for example, by preserving the better source for a particular database combination or by aiding full-text retrieval from a preferred source.

*Text Normalization*: It enhances deduplication by standardizing textual inputs before comparison. This includes converting text to lowercase and ignoring punctuation, which mitigates variations resulting from formatting differences, typographical errors, or inconsistent data entry. By reducing discrepancies in titles and author names, normalization improves both sensitivity (detecting true duplicates) and specificity (avoiding false positives). Given these benefits, normalization is enabled by default for title-based matching in most deduplication workflows, consistent with standard practice in citation deduplication. Researchers working with datasets containing non-Latin scripts or extensively abbreviated metadata should



XSL•FO
RenderX

verify that normalization performs as expected in their specific dataset before proceeding with full-scale resolution.

*Customization Protection*: It protects any articles with customized actions, such as assigned decisions, reasons, or labels, from being accidentally deleted during autoresolution (default: on). Customization Protection logic also checks if the user has taken a manual deduplication decision to keep the article. This feature is beneficial when importing new records after a review has already started.

## Discussion

### Overview

This discussion provides an overview of Rayyan's integrated deduplication capabilities, as previously reported [10], but with the enhanced feature of the Systematic Auto Resolver to support the Rayyan Method. While several software platforms offer deduplication features, many rely on fully automated algorithms with limited user control [5,7,13]. These tools may be efficient but can lack the flexibility needed to address nuanced deduplication scenarios, such as the problem posed by dirty data, including missing, incomplete, and nonstandardized metadata. Rayyan's Systematic Auto Resolver combines the strengths of more sensitive detection with automation and user control, allowing review teams to tailor the deduplication process to their specific needs.

### Understanding the Systematic Auto Resolver Performance in Deduplication

Unlike fully automated deduplication tools that apply fixed, predetermined algorithms without user input, the Systematic Auto Resolver executes only the criteria explicitly selected and authorized by the research team. While Rayyan provides recommended default settings to guide users toward established best practices, the criteria driving duplicate resolution are not fixed; instead, they are configurable, transparent, and documented at the user's discretion. As a result, the Rayyan Method for Deduplication, including the Systematic Auto Resolver, reflects the stringency and appropriateness of the criteria determined by the research team rather than a black-box algorithm. This design ensures that deduplication standards align with each review's specific requirements and that the transparent reporting of user-applied criteria provides for the auditability and reproducibility of the deduplication method used.

### Methodological Advantages

With its simple 2-stage process, the Rayyan Method can identify potential duplicates across datasets containing hundreds of thousands of references, leveraging its cloud-based infrastructure to handle large-scale work efficiently. In the detection stage, Rayyan produces a list of possible duplicates, each assigned a similarity score and presented in a clear table view, allowing researchers to assess which records may be duplicates and proceed to the resolution stage.

The resolution stage offers extensive customization, setting it apart from other tools. Researchers can view and manually review suspected duplicates, making decisions based on their specific criteria. They can also apply normalization rules, such as converting text to lowercase or omitting punctuation, to improve matching accuracy. This is particularly valuable when consolidating references exported from multiple databases, where differences in formatting conventions (eg, title capitalization, punctuation style, or author name formatting) can cause true duplicates to appear as distinct records. By applying normalization rules prior to comparison, researchers reduce the likelihood that format-driven inconsistencies across sources will result in missed duplicates or inflated reference counts.

Additionally, users can consider multiplicity by examining publication types or correcting previous record discrepancies. Rayyan's ability to allow for the combination of systematic rule application and manual review enhances accuracy while avoiding the cumbersome, rigid workflows seen in some manual methods or the concerns of false positives of other automated methods. Research teams may apply criteria sequentially, reviewing results after each pass. This allows teams to begin with conservative rules, observe their impact on resolving possible duplicates, and then decide whether to proceed with manual resolution or adjust to a different set of criteria. Ultimately, this approach maximizes both performance and convenience, accommodating the unique characteristics of each dataset and ensuring transparent, reproducible results while preserving ultimate decision-making for the reviewer.

### Transparency and Auditability

This tutorial intends not only to provide a step-by-step guide to the enhanced deduplication process in Rayyan but also to highlight the benefits of this feature to review teams. It is important to distinguish between the two stages of the Rayyan Method: duplicate detection, which uses an algorithm to identify and score possible duplicates, and duplicate resolution, which is entirely determined by the end user. While the detection algorithm automates the identification of candidate duplicates, all resolution decisions, including the criteria applied and the final determination for each record, are made exclusively through human control and oversight. This design does not eliminate automation from the deduplication process but ensures that automation is confined to candidate identification rather than final decision-making.

This distinction is critically impactful to 2 overarching objectives of the Rayyan platform: (1) reducing the risk of introducing bias through opaque automated resolution decisions, what we term vendor bias, while acknowledging that user-controlled resolution is only as rigorous as the methodological competency of the research team applying it and (2) providing a reproducible, auditable process that upholds the foundation of systematic review methodology. Transparency and auditability are critical in systematic reviews, particularly in regulatory contexts. Rayyan facilitates this by providing a clear record of all resolution decisions, including the user-specified criteria applied to the Systematic Auto Resolver and the rationale for resolving individual duplicates. This level of documentation is intended to enhance the credibility and defensibility of the review's methodological process.



XSL•FO
RenderX

### Limitations and Considerations

While the Rayyan Method provides significant control, this flexibility requires careful consideration by users. Incorrectly configured criteria could result in either false negatives (failing to identify true duplicates) or false positives (incorrectly identifying distinct records as duplicates), limited to the detected candidates. For example, deduplicating by DOI alone is theoretically feasible given the reasons for DOIs in the first place. However, the amount of erroneous, missing, and shared DOIs increases the possibility that this singular deduplication criterion alone can lead to false positives or false negatives. Adding more criteria increases accuracy, while reducing criteria increases sensitivity. Rayyan is under continuous development. Future development of the Systematic Auto Resolver is expected to continuously improve the accuracy, sensitivity, and usability of this capability across diverse use cases. Feedback from the research community serves as a major source of inspiration to address real-world contexts and identify the types of duplicates that are most challenging to resolve.

## Conclusion

Rayyan's duplicate detection and Systematic Auto Resolver features offer a unique combination of user control, flexibility, and support for transparent workflows. By empowering review teams to tailor the deduplication process to their specific needs, Rayyan contributes to more accurate, reproducible, and efficient systematic reviews. While this manuscript provides a descriptive overview of Rayyan's deduplication features, further empirical research will help guide future development. We encourage review teams to leverage Rayyan's deduplication features to enhance the rigor and efficiency of their systematic reviews while also contributing to the growing body of evidence on the effectiveness of different deduplication methods.

### Acknowledgments

The authors would like to acknowledge all members of the Rayyan team who contributed to the development of the platform features. During the finalization of the last submitted revision, the authors used Claude Sonnet 4.6 to review the manuscript for readability, language, and grammar. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

### Funding

No external funding was provided to support this work.

### Data Availability

All data and materials considered in this review are publicly available on the Rayyan website.

### Authors' Contributions

MMK conceived the study. MMK, AMS, HH, MA, and RA collected and analyzed software platform data and components. MMK, AMS, and RA defined all components of the manuscript and co-wrote the manuscript. All authors have read and approved the final manuscript. The corresponding author attests that the listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

### Conflicts of Interest

All authors (MMK, AMS, HH, MA, and RA) have completed the ICMJE Disclosure Form to disclose that they are employees of Rayyan Systems, Inc, the company that develops Rayyan. AMS, HH, MA, and RA are shareholders of Rayyan Systems, Inc.

### References

1. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
2. Whitlock EP, Lopez SA, Chang S, Helfand M, Eder M, Identifying FN. Identifying, selecting, and refining topics. In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews: Agency for Healthcare Research and Quality (AHRQ); 2008. URL: https://www.ncbi.nlm.nih.gov/books/NBK47096/ [accessed 2026-06-24]
3. Methods Guide for Effectiveness and Comparative Effectiveness Reviews: Agency for Healthcare Research and Quality (AHRQ); 2008. URL: https://www.ncbi.nlm.nih.gov/books/NBK47095/ [accessed 2026-06-24]
4. Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. J Med Libr Assoc 2016 Jul;104(3):240-243. [doi: 10.3163/1536-5050.104.3.014] [Medline: 27366130]
5. Hair K, Bahor Z, Macleod M, Liao J, Sena ES. The Automated Systematic Search Deduplicator (ASySD): a rapid, open-source, interoperable tool to remove duplicate citations in biomedical systematic reviews. BMC Biol 2023 Sep 7;21(1):189. [doi: 10.1186/s12915-023-01686-z] [Medline: 37674179]
6. Forbes C, Greenwood H, Carter M, Clark J. Automation of duplicate record detection for systematic reviews: deduplicator. Syst Rev 2024 Aug 2;13(1):206. [doi: 10.1186/s13643-024-02619-9] [Medline: 39095913]



XSL•FO
RenderX

7. McKeown S, Mir ZM. Considerations for conducting systematic reviews: evaluating the performance of different methods for de-duplicating references. Syst Rev 2021 Jan 23;10(1):38. [doi: 10.1186/s13643-021-01583-y] [Medline: 33485394]
8. Rathbone J, Carter M, Hoffmann T, Glasziou P. Better duplicate detection for systematic reviewers: evaluation of Systematic Review Assistant-Deduplication Module. Syst Rev 2015 Jan 14;4(1):6. [doi: 10.1186/2046-4053-4-6] [Medline: 25588387]
9. Typology of duplicate records in systematic review context. Medium. URL: https://farhadinfo.medium.com/typology-of-duplicate-records-in-systematic-review-context-97d2d94908b6 [accessed 2026-06-24]
10. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210. [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
11. Mosqueira-Rey E, Hernández-Pereira E, Alonso-Ríos D, Bobes-Bascarán J, Fernández-Leal Á. Human-in-the-loop machine learning: a state of the art. Artif Intell Rev 2023 Apr;56(4):3005-3054. [doi: 10.1007/s10462-022-10246-w]
12. Scott AM, Forbes C, Clark J, Carter M, Glasziou P, Munn Z. Systematic review automation tools improve efficiency but lack of knowledge impedes their adoption: a survey. J Clin Epidemiol 2021 Oct;138:80-94. [doi: 10.1016/j.jclinepi.2021.06.030] [Medline: 34242757]
13. Guimarães NS, Ferreira AJF, Ribeiro Silva RC, et al. Deduplicating records in systematic reviews: there are free, accurate automated ways to do so. J Clin Epidemiol 2022 Dec;152:110-115. [doi: 10.1016/j.jclinepi.2022.10.009] [Medline: 36241035]
14. Janka H, Metzendorf MI. High precision but variable recall – comparing the performance of five deduplication tools. J Eur Assoc Health Info Libr 2024;20(1):12-17. [doi: 10.32384/jeahil20607]
15. Hammer B, Virgili E, Bilotta F. Evidence-based literature review: de-duplication a cornerstone for quality. World J Methodol 2023 Dec 20;13(5):390-398. [doi: 10.5662/wjm.v13.i5.390] [Medline: 38229943]


## Abbreviations

**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses


*Edited by M Balcarras; submitted 27.Feb.2026; peer-reviewed by M Ansell, T Heck; revised version received 10.Jun.2026; accepted 10.Jun.2026; published 14.Jul.2026.*

*Please cite as:*
*Kim MM, Shanaa AM, Hammady H, Aboelnour M, Ayan R*
*Enhancing Transparency, Auditability, and Reproducibility of Deduplication in Systematic Reviews: Tutorial for the Rayyan Method and Systematic Auto Resolver*
*Interact J Med Res 2026;15:e94317*
*URL: https://www.i-jmr.org/2026/1/e94317*
*doi:10.2196/94317*


© Mimi M Kim, Amin M Shanaa, Hossam Hammady, Mohammad Aboelnour, Robert Ayan. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 14.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Real-World Clinical Characterization of Major Depressive Disorder and Treatment-Resistant Depression Supported by Natural Language Processing: Multicenter Observational Study From the MOOD Project

Dieter Zeeuws[1,2], MD; Katrien Bernagie[3], MD; Geert De Bruecker[3], MD; Isabel Claeys[4], MD; Charlotte Evenepoel[5], PhD; Fabienne Ver Donck[5], PhD; David Smeets[6], MS; Elke Peeters[7], MD

[1]Department of Psychiatry, Universitair Ziekenhuis Brussel, Jette, Brussels Capital, Belgium

[2]Neuroprotection and Neuromodulation Research Group (NEUR), Center for Neurosciences (C4N), Vrije Universiteit Brussel, Brussels, Brussels Capital, Belgium

[3]Psychiatric Center Ariadne, Lede, Belgium

[4]AZ Maria Middelares, Ghent, Belgium

[5]LynxCare Clinical Informatics, Leuven, Belgium

[6]Department External and Government Affairs, Johnson & Johnson, Beerse, Belgium

[7]Department Medical Affairs, Johnson & Johnson, Beerse, Belgium

**Corresponding Author:**
Dieter Zeeuws, MD
Department of Psychiatry
Universitair Ziekenhuis Brussel
Laarbeeklaan 103
Jette, Brussels Capital, 1090
Belgium
Phone: 32 024774111
Email: dieter.zeeuws@uzbrussel.be

## Abstract

**Background:** Major depressive disorder (MDD) and treatment-resistant depression (TRD) are heterogeneous conditions in which key clinical details are split across structured fields and free-text notes in electronic health records (EHRs), constraining population-level insight and timely audit of care quality.

**Objective:** This study aims to present a clinician-oriented, artificial intelligence-supported real-world evidence (RWE) methodology integrating structured and unstructured EHR data to profile MDD and TRD, and report comorbidity patterns from a 2-site pilot. This analysis reports the first objective of the MOOD project, which is to characterize the real-world clinical and disease severity profile of patients with MDD and treatment-resistant depression, providing a necessary foundation for subsequent evaluations of treatment patterns and outcomes.

**Methods:** We conducted a retrospective study in 2 Belgian hospitals (September 2021-June 2023). Adults (aged ≥18 years) with MDD were identified via *DSM-IV* (*Diagnostic and Statistical Manual of Mental Disorders* [Fourth Edition]) and *ICD-10* (*International Statistical Classification of Diseases, Tenth Revision*) codes or natural language processing-detected note mentions; bipolar depression was excluded. TRD was defined as initiation of a third distinct antidepressant, supplemented by explicit mentions of TRD in notes. Structured data (demographics, diagnoses, medications, and hospitalizations) were harmonized in an Observational Medical Outcomes Partnership warehouse. Free-text notes were processed with a natural language processing pipeline to capture symptoms, psychiatric comorbidities, and contextual events.

**Results:** We identified 1147 adults with MDD, of which 46% (524/1147) met TRD criteria. Females comprised 62.9% (722/1147) and mean (SD) age was 57.8 (18.4) years. Mortality was 13.3% (152/1147) overall (57/1147, 10.9% TRD vs 95/1147, 15.2% non-TRD). Common medical comorbidities were central nervous system diseases (477/1147, 41.6%) and heart diseases (349/1147, 30.4%). Dementia was more frequent in TRD (42/1147, 8% vs 32/1147, 5.1%), whereas obesity was higher in non-TRD (70/1147, 11.2% vs 46/1147, 8.8%). Anxiety disorder occurred in 35.4% (406/1147) overall and was more prevalent in TRD (229/1147,



XSL•FO
RenderX

43.7% vs 177/1147, 28.4%); personality and panic disorders also trended higher. Severity was sparsely documented (severe MDD 170/1147, 14.8%) and standardized scales were rarely recorded.

**Conclusions:** We present a step-by-step artificial intelligence-supported methodology tailored for clinicians, discussing challenges in integrating RWE into psychiatry, and identifying opportunities to enhance data collection with minimal workflow changes, which emphasizes the transformative potential of RWE systems in mental health research.

***(Interact J Med Res 2026;15:e86448)*** doi:10.2196/86448

**KEYWORDS**

major depressive disorder; treatment-resistant depression; real-world evidence; natural language processing; symptom clusters; longitudinal data; clinical practice

## Introduction

Major depressive disorder (MDD) affects a considerable portion of the global population, often leading to significant disability [1]. Meanwhile, treatment-resistant depression (TRD), where patients do not respond adequately to at least 2 standard antidepressant treatments, remains an especially difficult clinical challenge [2]. Studies have shown that patients with TRD often face poorer outcomes, which amplifies the need for tailored therapeutic approaches [2]. Beyond treatment nonresponse, TRD is also characterized by substantial clinical heterogeneity (eg, comorbidity profiles, symptom expression, and illness course), which complicates comparisons across cohorts and health systems and makes baseline characterization clinically relevant in its own right [3]. Large-scale electronic health record (EHR)—based studies have previously used TRD as a model phenotype and highlighted the need for reliable approaches to define outcomes and longitudinal status in routine care [4]. The MOOD project focuses on a comprehensive description of this patient population, aiming to elucidate the demographics, medical history, and psychiatric patient history among MDD and TRD cohorts.

Real-world evidence (RWE) has become increasingly important to complement or expand upon findings from controlled trial settings and fill data gaps on new treatments, for example, patient profiling [5]. Furthermore, natural language processing (NLP) offers a time-efficient means to mine free-text fields, which are abundant in psychiatric reports [6]. This is particularly relevant in mental health EHRs because clinically decisive information such as symptom clusters, course specifiers, contextual triggers, and clinician-assessed status frequently appears in narrative documentation rather than in standardized structured fields [4,6,7]. Moreover, psychiatric research is hindered by a lack of standardized documentation, particularly for severity levels and outcomes, which are often inconsistently captured or absent in structured data [8-10]. Empirical work has also shown that even when severity indicators exist (eg, patient health questionnaires), their availability can be incomplete or uneven across encounters [11], reinforcing the need to combine structured fields with information from clinical text when feasible. This work showcases the methodology of a pilot project in psychiatry aimed at streamlining the comprehensive analysis of MDD and TRD patient populations.

We present a detailed methodology of a preliminary study conducted at 2 Belgian hospitals. The study leverages NLP as an artificial intelligence approach to analyze unstructured data alongside structured data from EHRs in psychiatry [12-15], generating RWE that seeks to offer a nuanced understanding of patient experiences and treatment trajectories even in the absence of depression scale scores. In doing so, the approach aligns with broader developments in mental health informatics, where NLP has been used to structure depression-related information from large EHR repositories to support observational research at scale [7]. Given the clinician-facing orientation of this manuscript, the NLP workflow is described step-by-step with emphasis on interpretability, validation, and how outputs map back to clinical concepts, consistent with clinician-oriented guidance on the development and implementation of AI methods in medicine [16]. Initial results on patient demographics and comorbidities are included here, and future analyses will address patterns of care. The ultimate goal is to inform clinical practice and guide future research into the complexities of treating depression in diverse patient cohorts.

The MOOD project comprises three overarching objectives focused on patients with MDD, with particular emphasis on those meeting criteria for TRD. The first objective is to characterize patient demographics, medical history, and psychiatric disease burden, including MDD severity, in these populations. The second objective is to assess real-world treatment patterns, encompassing pharmacological therapies, psychotherapy, and neuromodulation approaches used in MDD and TRD. The third objective aims to address data gaps for emerging treatments, with a specific focus on treatment patterns associated with esketamine nasal spray in patients with TRD [3]. While the MOOD project encompasses multiple objectives, the comprehensive real-world characterization of patients with MDD and treatment-resistant depression represents an important scientific objective in its own right. Despite the clinical relevance of TRD, contemporary data describing disease severity, comorbidity burden, and longitudinal clinical profiles in routine care remain limited [8-10]. The present manuscript addresses this gap by reporting the results of the first project objective, thereby establishing a baseline understanding of the studied population prior to analysis of treatment patterns and outcomes, which are ongoing.

## Methods

### Study Design and Data Sources

This study was a preliminary multicenter analysis conducted at 2 Belgian hospitals (University Hospital Brussels and Psychiatric Center Ariadne) from September 2021 to June 2023. Data were

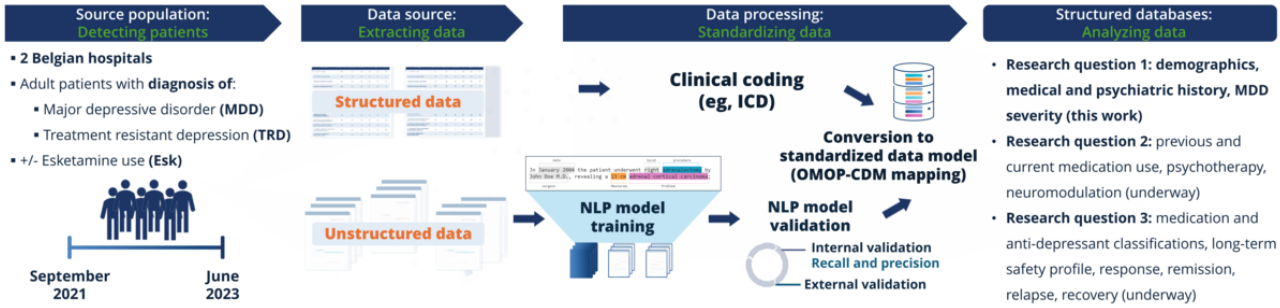


XSL•FO
RenderX

extracted from both structured and unstructured sources within the EHRs of the hospital. Structured data encompasses demographic details, treatment dosages, frequency of administration, and hospitalization information, whereas unstructured data comprised clinicians' notes, pathology reports, and other textual documentation that captured patient histories and treatment responses.

The integration of data sources facilitated a thorough analysis of patient characteristics and enabled the study to capture information frequently overlooked in traditional research settings, such as nuanced clinical observations noted only in free-text fields. As the available data originate from two Belgian hospitals, the findings reflect the clinical documentation structure of these settings, with potential for broader generalizability as additional sites are incorporated in future work. A schematic of the overall study design is provided in Figure 1.

**Figure 1.** Schematic of the study design and methodology. Results for Research question one are included in this work, while analyses for Research questions 2 and 3 are underway. Esk: esketamine; *ICD*: *International Classification of Diseases*; MDD: major depressive disorder; NLP: natural language processing; OMOP-CDM: Observational Medical Outcomes Partnership-common data model; TRD: treatment resistant depression.

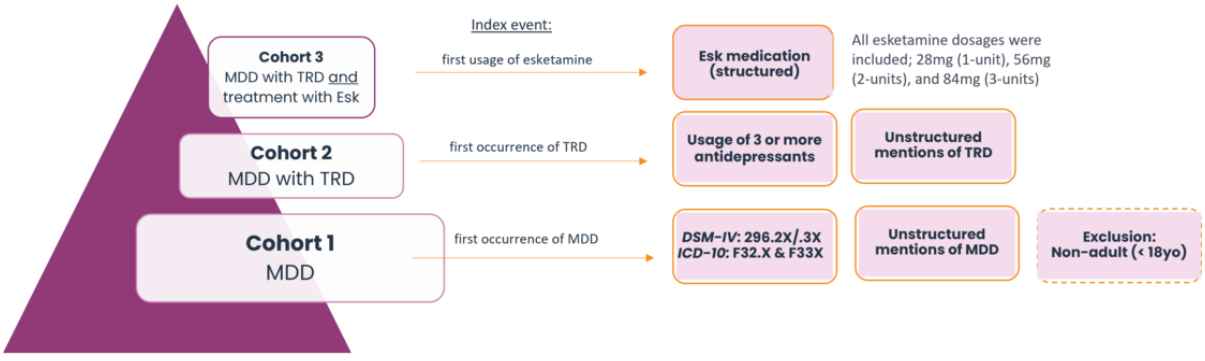



## Participants and Cohorts

Patients eligible for inclusion were those aged 18 years or older with any record of MDD or TRD, and some had documented esketamine (Esk) use. Treatment-resistant depression is defined by the European Medicines Agency (EMA) as MDD with inadequate response to at least 2 different antidepressant treatments administered at adequate dose and duration [17]. Likewise, EMA approved esketamine specifically for adults with TRD [18], such that patients with TRD were assumed to have a diagnosis of MDD, and patients with esketamine use were assumed to have both MDD and TRD. This operational definition prioritizes sensitive real-world identification of TRD in EHR data; reasons for treatment changes (eg, nonresponse vs intolerance) are not consistently documented in structured records and could not be reliably distinguished. This definition is also consistent with prior large-scale TRD studies such as STAR*D [19], which similarly rely on sequencing of antidepressant trials in the absence of systematically documented reasons for change [20-22].

Three nested cohorts were established (Figure 2), including a MDD cohort (diagnosed through *ICD-10* [*International Statistical Classification of Diseases, Tenth Revision*] codes F32.X or F33.X, as well as *DSM-IV* [*Diagnostic and Statistical Manual of Mental Disorders* {Fourth Edition}] and *DSM-5* [*Diagnostic and Statistical Manual of Mental Disorders* {Fifth Edition}] codes 296.2X, 296.3X, excluding bipolar depression), a TRD cohort (individuals with three or more distinct antidepressants during the study period, or explicit mention of TRD in unstructured notes), and an esketamine (Esk) cohort (patients with esketamine treatment events of any dose). The index date for the MDD cohort was defined as the first occurrence of either MDD, TRD, or Esk, for the TRD cohort as the first occurrence of either TRD or Esk, and for the Esk cohort as the first recorded esketamine administration. Maximum follow-up (censoring) was applied based on the latest recorded date for any event per patient, defined as any recorded date for medication, diagnosis, or treatment in both structured and unstructured data.

**Figure 2.** Schematic of the nested cohorts included in the study, their index events, and definitions. *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*; Esk: esketamine; *ICD-10*: *International Classification of Diseases*, Tenth Revision; MDD: major depressive disorder; TRD: treatment resistant depression.






XSL•FO
RenderX

## Variables

The variables were grouped according to the MOOD study's primary objectives, with corresponding timeframes and patient cohorts of interest (Table 1). This study presents the results for demographics, medical history, and psychiatric history, which constitute the first objective (shown as "this work" in Table 1). Future analyses will cover the additional variables linked to pharmacological and psychotherapeutic interventions, esketamine treatment patterns, and long-term outcomes (shown as "ongoing" in Table 1). Specific medications and antidepressant classifications can be found in Table S1 in Multimedia Appendix 1.

**Table 1.** Variables used in the study per objective, with their respective cohorts and timeframe.

| Objective | Variables extracted | Cohorts | Timeframe |
|---|---|---|---|
| Patient demographics, medical, and psychiatric history (this work) | • Demographics: age, sex, death status.<br>• Medical history: Parkinson disease, obesity, cardiac arrhythmia, rheumatoid arthritis, kidney disease, COPD[a], stomach disease, heart disease, anemia, central nervous system diseases, dementia, liver disease, HIV, diabetes, thyroid disease, cancer, stroke, CVS[b] or fibromyalgia, hypercholesterolemia, autoimmune disease, infectious diseases, and hypertension.<br>• Psychiatric history: borderline personality disorder, bulimia nervosa, obsessive-compulsive disorder, circadian rhythm disorders, anxiety disorder, addiction (composite), personality disorder, panic disorder, posttraumatic stress disorder, anorexia nervosa, suicidal tendencies.<br>• MDD[c] severity: mild, moderate, and severe MDD, MADRS[d], CGI-S[e]. | • MDD and TRD[f] (this work), Esk[g] (ongoing) | • On or before index date |
| Medication, psychotherapy, and neuromodulation treatments (ongoing) | • Medication use: SSRIs[h], tricyclic antidepressants, atypical antipsychotic augmentation, serotonin and noradrenaline reuptake inhibitors, monoamine oxidase inhibitors, anxiolytics, selective noradrenaline reuptake inhibitors, antidepressants acting on neuroreceptors, noradrenaline and dopamine reuptake inhibitors, other antidepressants.<br>• Psychotherapy: CBASP[i], cognitive behavioral therapy, psychoanalytic therapy, crisis intervention, psychodynamic psychotherapy, problem-solving therapy, mindfulness, relationship or family therapy, interpersonal psychotherapy, general psychotherapy.<br>• Neuromodulation: transcranial magnetic stimulation and electroconvulsive therapy. | • TRD and Esk | • On or before index date |
| Esketamine treatment patterns (ongoing) | • Treatment characteristics: Frequency, duration, dosing, switching, and cessation.<br>• Efficacy: Response, remission, relapse, recovery.<br>• Safety: Adverse events.<br>• Psychiatric hospitalization. | • Esk | • Within esketamine treatment period |

[a]COPD: chronic obstructive pulmonary disorder.

[b]CVS: cyclic vomiting syndrome.

[c]MDD: major depressive disorder.

[d]MADRS: Montgomery-Åsberg Depression Rating Scale.

[e]CGI-S: Clinical Global Impression - Severity Scale.

[f]TRD: treatment-resistant depression.

[g]Esk: esketamine.

[h]SSRI: selective serotonin reuptake inhibitor.

[i]CBASP: cognitive behavioral analysis system of psychotherapy.

## Database Generation

Following data extraction, the processed information was systematically integrated into a research-grade Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) data warehouse, which promotes consistent data storage and reproducibility [23]. Data mapping used the *ICD* (*International Classification of Diseases*) codes for structured data and their equivalent Unified Medical Language System (UMLS) Concept Unique Identifiers (CUIs), while unstructured



XSL•FO
RenderX

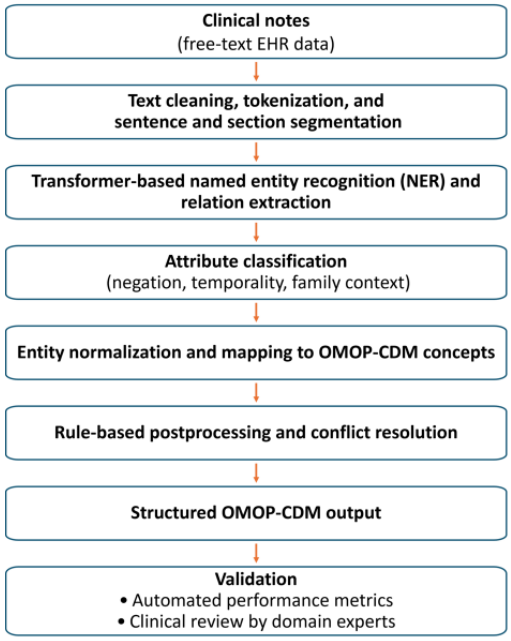
data were directly encoded to CUIs (see full data point list in Table S2 in Multimedia Appendix 1).

NLP pipeline as an initial step, the underlying clinical notes were characterized by determining the total number of records, the number of patients with at least one record, and the allocation of documents to the predefined note-type categories. These data informed the preprocessing stages of the NLP pipeline.

NLP was used to handle unstructured text derived from clinical notes, pathology reports, and other sources. In this pipeline, text preprocessing involved splitting documents into sentences and words, followed by entity recognition to identify clinical terms such as “remission” or “relapse.” Entity linking was performed to map recognized entities to standardized medical concepts in the UMLS. An attribute extraction phase was then used to capture further details such as negation (eg, “no suicidal ideation”) or speculation. Model training and fine-tuning leveraged extensive medical datasets, with adjustments made for the psychiatry context, as shown for other therapeutic areas in detail (see [12-15,24,25]).

Figure 3 provides an overview of the end-to-end workflow, from free-text EHR documentation to structured, analysis-ready OMOP-CDM outputs, and situates each NLP component within the overall data-generation process. In line with the accepted methodological description of our platform and prior validated implementations in other disease areas [12-15,24-26], the pipeline was implemented as a cascade of transformer-based models complemented by rule-based interventions defined by domain experts to support robust extraction in real-world clinical text.

**Figure 3.** Overview of the clinical natural language processing pipeline used to convert free-text electronic health record documentation into structured Observational Medical Outcomes Partnership-common data model outputs. The workflow begins with clinical notes and applies text cleaning and segmentation, transformer-based named entity recognition and relation extraction, and attribute classification (eg, negation, temporality, and family context). Extracted mentions are then normalized to standardized Observational Medical Outcomes Partnership-common data model concepts, followed by rule-based post-processing and conflict resolution. The resulting structured Observational Medical Outcomes Partnership-common data model dataset is evaluated using automated performance metrics and clinical review by domain experts. EHR: electronic health record; NER: named entity recognition; OMOP-CDM: Observational Medical Outcomes Partnership-common data model.




Text cleaning, tokenization, and sentence and section segmentation were first applied to standardize heterogeneous document formats and to support downstream extraction within meaningful clinical contexts (Figure 3). Transformer-based named entity recognition (NER) and relation extraction were then applied to identify clinically relevant entities (eg, symptoms, diagnoses, and medications) and their relations (eg, date or dosage). In the pipeline, NER and relation extraction are framed as a sequence-to-sequence structured prediction task, which generates a structured representation of entities and relations directly from raw text [25,26]. This approach supports joint extraction of entities and relations in a single step and allows nested or compositional mentions to be represented in the model output.

After entity extraction, attribute classification was used to capture the contextual status of each extracted mention, including negation, temporality, and family context (Figure 3). This step distinguishes affirmed findings from negated or uncertain statements (eg, “no suicidal ideation,” “possible relapse,” “history of depression,” or “family history of bipolar disorder”), thereby reducing false positives that can arise when narrative documentation contains ruled-out symptoms or historical information. The attribute model is fine-tuned for multiple Boolean or contextual attributes, enabling classification of mentions into categories such as absent or negated, hypothetical, historical, or family-related [25,26].

Entity normalization and mapping to standardized OMOP-CDM concepts were subsequently performed (Figure 3). Structured data elements retained their source coding (eg, *ICD* where



XSL•FO
RenderX

applicable) and were aligned to standardized concept identifiers, while extracted unstructured mentions were normalized to UMLS-linked concepts and mapped to the OMOP-CDM vocabulary. Normalization is implemented via transformer-based entity linking (eg, SapBERT-style metric learning over UMLS) with nearest-neighbor retrieval and threshold optimization; concept mapping covers relevant clinical vocabularies used within OMOP (eg, SNOMED CT [Systematized Nomenclature of Medicine – Clinical Terms], *ICD-10*, Logical Observation Identifiers Names and Codes [LOINC], and RxNorm) [25,26].

Finally, rule-based postprocessing and conflict resolution were applied to improve data fidelity and address recurrent real-world documentation artifacts (Figure 3). These interventions include pattern-based extraction for semistructured mentions, exclusion of nonclinical document sections, filtering of incompatible NER label-semantic type combinations, resolution of conflicting attributes (eg, historical labels paired with inconsistent dates), and default date assignment when timing is not explicitly documented, using a predefined cascade of rules [25,26].

### Training Data and Annotation Framework

The NLP models, used for initial processing of the corpus, rely on a continuously expanding proprietary corpus of manually annotated clinical text in Dutch. This annotated resource includes approximately 33897 sentence-level annotations across 5954 sentences (for NER, relation extraction, and attribute detection) and 4716 concept-level annotations from 3777 documents (for coding and attribute classification). For this study, model performance was estimated using a separate set of 2638 clinician-validated annotations derived from 1921 documents, which serve as an out-of-the-box (OOTB) evaluation sample. These validation annotations were subsequently reintegrated into the training pool as part of an iterative continuous-learning process, in which corrected model outputs inform further fine-tuning. This workflow reflects our standard development approach, prioritizing broad clinical coverage and iterative refinement rather than fixed, project-specific data partitions.

### Validation

Validation involved a multistep process to confirm data quality and clinical reliability (Figure 3). First, automated evaluation was performed by comparing NLP-extracted concepts against a clinician-curated gold standard, with precision, recall, and $F_1$ scores calculated at the datapoint level; where relevant, performance was also assessed separately for specific subtasks such as attribute handling (eg, negation and temporality) and concept normalization. Second, clinical review by domain experts was conducted on sampled patient records and on aggregated outputs to confirm that extracted concepts were clinically meaningful, internally consistent, and aligned with expectations from routine practice. This review step explicitly targeted common failure modes in psychiatric documentation, including ambiguous phrasing, historical versus current symptom mentions, and "rule-out" statements, and it served as an iterative quality check to identify systematic extraction errors that could affect downstream analyses. Findings from automated metrics and expert review informed targeted refinements of extraction logic and postprocessing (eg, resolving conflicting attributes or excluding nonclinical document sections), followed by reassessment to ensure that updates improved fidelity without introducing new inconsistencies [25,26].

### Statistical Analysis

Descriptive statistics were used to summarize findings related to the first objective (patient demographics, medical history, and psychiatric history). Means, medians, SDs, and minimum-maximum values were used for continuous variables such as age at index, whereas categorical variables, including comorbidity status, were presented as frequencies and percentages. Statistical analyses were performed using R environment software (version 4.1.0; R Foundation for Statistical Computing). No inferential analyses were included because the objective was descriptive cohort characterization rather than hypothesis testing. Counts below five were suppressed as "<5" (<x%), and when this alone still allowed back-calculation, the complementary subgroup was shown as a symmetric count and percentage range.

### Ethical Considerations

Ethical approval was obtained from the relevant Institutional Review Boards (Commissie Medische Ethiek, Universitair Ziekenhuis Brussel, BUN1432023000167, EC-2023-199, Project ID=23162_RWE Mood). The ethics evaluation resulted in approval for the secondary use of routinely collected health data. The study was conducted following the Declaration of Helsinki, ethical standards of the responsible institutional and national committees on human experimentation, and applicable regulations. Patient privacy was safeguarded according to Belgian Privacy Law and General Data Protection Regulation articles 6, 9, and 89. Data were pseudonymized prior to analysis. Patient consent was waived by the approving ethics committees due to pseudonymization of the data and the retrospective nature of the study, in accordance with Belgian legislation. No financial or other compensation was provided to patients.

## Results

### NLP Pipeline and Metrics

A total of 56,396 records across both hospitals were included in the dataset, where 1262 patients presented at least one record. Clinical assessment notes constituted the majority of all extracted records (32,075/56,396, 56.9%). Admission, intake, and hospitalization notes (13,109/56,396, 23.2%), together with consultation notes (5605/56,396, 9.9%, including client, referrer, bedside, psychiatric, and telephone contacts), accounted for most of the remaining documentation. The other note types, including multidisciplinary meeting reports, administrative or system-generated communications, diagnostic or procedural documents, risk and safety assessments, and treatment-plan notes, each represented a smaller proportion of the dataset (< 4% each).

Performance against a clinician-annotated gold standard, reported as precision and recall for OOTB pipeline versus the postmitigation pipeline (ie, after rule-based postprocessing and conflict resolution), is summarized in Table S3 in Multimedia Appendix 1. Overall, postmitigation precision was high, with values ranging from 66.7% to 100% (96.6% mean), while



XSL•FO
RenderX

postmitigation recall ranged from 91.7% to 100% (98.1% mean). In contrast, OOTB recall was notably more variable (range 10%-100%, 72.3% mean), supporting the importance of the mitigation layer for reliable extraction.

Performance gains were most visible for clinically central symptom and outcome concepts that are frequently documented in narrative form. For example, recall improved from 39.3% to 95.2% for depressed mood, from 42.4% to 93.9% for anxiety, and from 35.4% to 95.8% for suicidal mentions. Similarly, the concept Major Depressive Disorder increased from 10% to 95% recall after mitigation, indicating that postprocessing materially reduced missed detections for key diagnostic constructs when expressed variably in free text (Table S3 in Multimedia Appendix 1).

Across variable classes, medication entities and medication-class groupings tended to perform best after mitigation (typically near-perfect precision and recall for commonly observed treatments or classes in the annotated sample), as did many diagnosis or comorbidity concepts (postmitigation precision typically in the high-90% range with recall approaching 100%). Symptom or context concepts also reached strong postmitigation recall overall, but they represented the main area where OOTB performance was lower, consistent with the greater linguistic variability of symptom documentation in psychiatric notes.

Routine documentation sometimes indicated different intended uses for the same medication, which is not consistently recoverable from structured medication records alone. For example, trazodone and benzodiazepines were frequently mentioned in the context of sleep rather than antidepressant or anxiolytic use, and this indication context was captured from unstructured notes by the NLP extraction layer.

## Patient Demographics

Table 2 summarizes key demographic characteristics of patients with MDD, stratified into not-treatment-resistant (Not TRD, 623/1,147, 54%) and treatment-resistant (TRD, 524/1,147, 46%) subgroups. Among patients with TRD, 2% (n=8) had an explicit TRD mention in clinical notes; the remaining patients met the TRD definition based on medication information (≥3 distinct antidepressants). Females comprise 62.9% (722/1,147) of the overall MDD population, with a higher proportion in the TRD group (349/524, 66.6%) compared to Not TRD (373/623, 59.9%). The mean age is 57.8 years, and the median age is slightly higher in the TRD group (60 years) than in not TRD (58 years). Most patients are alive (995/1,147, 86.7%), with a slightly higher mortality rate in the Not TRD group (95/623, 15.2%) compared to TRD (57/524, 10.9%).

**Table 2.** Patient demographics.

| Characteristics | MDD[a] (Overall), n (%) | MDD (N=1147; 100%) | |
|---|---|---|---|
| | | Not TRD[b], n (%) (N=623; 54% of MDD) | TRD, n (%) (N=524; 46% of MDD) |
| **Sex, n (%)** | | | |
| Female | 722 (62.90) | 373 (59.90) | 349 (66.60) |
| Male | 425 (37.10) | 250 (40.10) | 175 (33.40) |
| **Age at index** | | | |
| Mean (SD) | 57.8 (18.4) | 57.2 (18.7) | 58.4 (18) |
| Median (min-max) | 59 (18-101) | 58 (19-96) | 60 (18-101) |
| **Status, n (%)** | | | |
| Alive | 995 (86.70) | 528 (84.80) | 467 (89.10) |
| Dead | 152 (13.30) | 95 (15.20) | 57 (10.90) |

[a]MDD: major depressive disorder.

[b]TRD: treatment-resistant depression.

## Medical and Psychiatric History of MDD and Patients With TRD

Table 3 summarizes the medical and psychiatric history of patients with MDD, with or without TRD. Among medical comorbidities, central nervous system diseases, infections, and heart diseases are the most common, with similar prevalence across both subgroups. Notable differences include higher rates of dementia in patients with TRD (42/524, 8% vs 32/623, 5.1%) and obesity in individuals without TRD (70/623, 11.2% vs 46/524, 8.8%). Psychiatric comorbidities reveal more pronounced differences, with anxiety disorders found at higher proportions for patients with TRD (229/524, 43.7% vs 177/623, 28.4%), along with higher rates of personality disorders and panic disorders. Alcohol and drug-related issues are prevalent in both groups, with slightly higher alcohol abuse (148/623, 23.8%) and drug dependence (152/623, 24.4%) in patients without TRD. In terms of depression severity, detected on or at the index date, severe MDD is observed in 14.8% (170/1147) of the overall population, with comparable rates across subgroups. Severity was only well captured in a minority of patients (see the Discussion section). This data underscores the complex clinical profiles of MDD and highlights key differences between treatment-resistant and nontreatment-resistant cases.



XSL•FO
RenderX

**Table 3.** Patient history.

| Characteristics | MDD[a] (Overall) (N=1147) | MDD (N=1147; 100%) | |
|---|---|---|---|
| | | Not TRD[b] (N=623; 54% of MDD) | TRD (N=524; 46% of MDD) |
| **Medical history, n (%)** | | | |
| Anemia | 175 (15.3) | 94 (15.1) | 81 (15.5) |
| Autoimmune disease | 33 (2.9) | 18 (2.9) | 15 (2.9) |
| Cancer | 116 (10.1) | 58 (9.3) | 58 (11.1) |
| Cardiac arrhythmia | 210 (18.3) | 112 (18) | 98 (18.7) |
| Central nervous system diseases | 477 (41.6) | 261 (41.9) | 216 (41.2) |
| COPD[c] | 100 (8.7) | 59 (9.5) | 41 (7.8) |
| CVS[d] or fibromyalgia | 38 (3.3) | 10 (1.6) | 28 (5.3) |
| Dementia | 74 (6.5) | 32 (5.1) | 42 (8) |
| Diabetes | 193 (16.8) | 121 (19.4) | 72 (13.7) |
| Heart diseases | 349 (30.4) | 190 (30.5) | 159 (30.3) |
| HIV | 11 (1) | 5 (0.8) | 6 (1.1) |
| Hypercholesterolemia | 196 (17.1) | 110 (17.7) | 86 (16.4) |
| Hypertension | 306 (26.7) | 169 (27.1) | 137 (26.1) |
| Infection | 344 (30) | 187 (30) | 157 (30) |
| Kidney disease | 181 (15.8) | 100 (16.1) | 81 (15.5) |
| Liver disease | 107 (9.3) | 60 (9.6) | 47 (9) |
| Obesity | 116 (10.1) | 70 (11.2) | 46 (8.8) |
| Parkinson | 23 (2) | 10 (1.6) | 13 (2.5) |
| Rheumatoid arthritis | 6 (0.5) | <5[e] | <5 |
| Stomach disease | 134 (11.7) | 65 (10.4) | 69 (13.2) |
| Stroke or cerebrovascular accident | 32 (2.8) | 15 (2.4) | 17 (3.2) |
| Thyroid disease | 153 (13.3) | 78 (12.5) | 75 (14.3) |
| **Psychiatric history, n (%)** | | | |
| Alcohol abuse | 250 (21.8) | 148 (23.8) | 102 (19.5) |
| Anorexia nervosa | 12 (1) | 5 (0.8) | 7 (1.3) |
| Anxiety disorder | 406 (35.4) | 177 (28.4) | 229 (43.7) |
| Bipolar disorder | 67 (5.8) | 30 (4.8) | 37 (7) |
| Borderline personality disorder | 43 (3.7) | 19 (3) | 24 (4.6) |
| Bulimia nervosa | <5 | <5 | <5 |
| Circadian rhythm disorders | <5 | <5 | <5 |
| Drug abuse | 85 (7.4) | 37 (5.9) | 48 (9.2) |
| Drug dependence | 254 (22.1) | 152 (24.4) | 102 (19.5) |
| Obsessive-compulsive disorder | <5 | <5 | <5 |
| Panic disorder | 13 (1.1) | <5 | 7-13 (1.3-2.5) |
| Personality disorder | 108 (9.4) | 44 (7.1) | 64 (12.2) |
| Posttraumatic stress disorder | 25 (2.2) | 13 (2.1) | 12 (2.3) |
| Suicidal | 334 (29.1) | 174 (27.9) | 160 (30.5) |
| Addiction | 298 (26) | 165 (26.5) | 133 (25.4) |



XSL•FO
RenderX

| Characteristics | MDD[a] (Overall) (N=1147) | MDD (N=1147; 100%) | |
|---|---|---|---|
| | | Not TRD[b] (N=623; 54% of MDD) | TRD (N=524; 46% of MDD) |
| CGI-S[f] | —[g] | — | — |
| MADRS[h] | — | — | — |
| Mild major depressive disorder | <5 | <5 | <5 |
| Moderate major depressive disorder | 22 (1.9) | 15-21 (2.4-3.4) | <5 |
| Severe major depressive disorder | 170 (14.8) | 96 (15.4) | 74 (14.1) |

[a]MDD: major depressive disorder.

[b]TRD: treatment-resistant depression.

[c]COPD: Chronic Obstructive Pulmonary Disease.

[d]CVS: Cyclic Vomiting Syndrome.

[e]These groups contains less than 5 patients; hence, percentages are not available for these groups. Precise numbers are masked per the contract between the hospitals and the sponsor.

[f]CGI-S: Clinical Global Impressions-Severity.

[g]Not available.

[h]MADRS: Montgomery-Åsberg Depression Rating Scale.

## Discussion

### Overview, Scope, and Importance of the MOOD Project

The MOOD project highlights the growing need for RWE in psychiatry, particularly for understanding complex disorders such as MDD and its treatment-resistant form [8,27]. The findings presented here provide an essential real-world baseline of disease burden and patient complexity in MDD and TRD, enabling more informed interpretation of subsequent treatment-focused analyses from this project. By integrating structured and unstructured data, the study addresses key gaps in psychiatric research, where conventional methods often overlook contextual details captured only in free-text notes [12,13]. This approach is vital given the global prevalence and significant disease burden of depression, as well as the unique clinical and economic challenges posed by TRD [2,3]. Building upon existing literature that highlights the limitations of traditional datasets, the MOOD project aims to generate comprehensive insights that can inform a more patient-centered and evidence-driven model of psychiatric care [5,6,8].

### Main Findings and Link to Previous Work

Results regarding NLP metrics support the contribution of unstructured text augmentation: variables that are often absent or inconsistently represented in structured fields (symptoms, clinical course descriptors, response language) can be extracted with high postmitigation performance, aligning with prior evidence that NLP improves ascertainment beyond structured claims data alone in large-scale psychiatric EHR studies [4]. The first analyses from the MOOD project reveal how NLP-based extraction of free-text data complements structured clinical fields, leading to a richer depiction of patient trajectories in MDD and TRD. The systematization of clinical experience results in a holistic overview that avoids confirmation bias based on individual cases. [6,7]. For instance, the project's NLP pipeline successfully extracted a dashboard overview of clinical experience in TRD.

Automation of data extraction through NLP further reduces the burden on clinicians, allowing them to spend more time with patients while still generating valuable research insights [6]. This aligns with the broader goal of enhancing both research and clinical practice in psychiatry. These findings align with prior large-scale investigations showing that real-world data capture a heterogeneous array of patient experiences and outcomes [8,27]. The capacity to identify these factors using textual data sources resonates with studies showing that TRD can have varying definitions and multistep strategies commonly are recommended, which may involve switching, augmentation, or combination regimens [3]. Consequently, the MOOD findings expand on this literature by illustrating how robust informatics frameworks can overcome the inconsistent or complex documentation required for psychiatric research.

As extensively discussed in major TRD reviews and large real-world cohorts [20-22], heterogeneity in TRD definitions and reliance on retrospective EHR documentation contribute to variability in reported prevalence; our findings therefore align with expectations for hospital-based MDD populations [22], where TRD proportions typically approach 45%-50% rather than the approximately 30% seen in general MDD populations [3]. In addition, the descriptive differences observed in our cohort, such as higher rates of dementia, anxiety disorders, personality disorders, and panic disorder among patients with TRD, are broadly consistent with patterns described in published European TRD cohorts, which similarly report greater psychiatric and medical complexity in TRD than non-TRD populations [28-30].

### Disentangling the Depression Symptom Cluster Using Unstructured Data

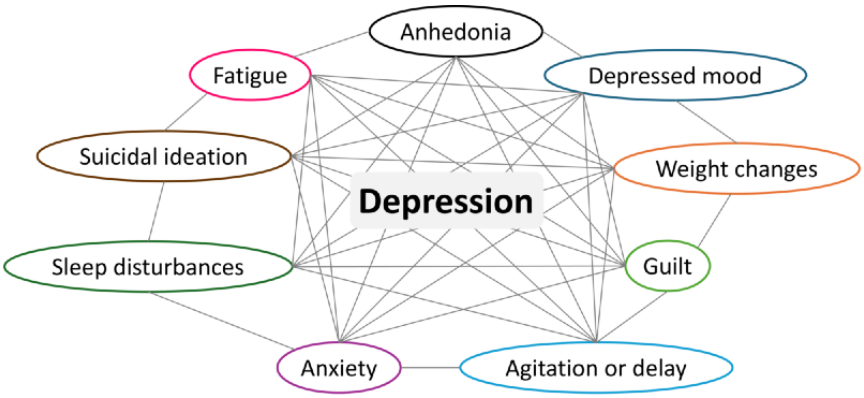
The study highlights the importance of a structured framework for understanding the "nine pillars of depression," or "depression



XSL•FO
RenderX

symptom cluster" that defines MDD in the *DSM-V* [31], which encompasses key domains such as diagnosis, symptom severity, and quality of life (Figure 4). While structured data sources offered a baseline for documenting diagnoses and comorbidities, unstructured data provided critical insights into the nuanced clinical profiles of patients. Indeed, patient profiles within MDD can vary widely depending on symptom patterns [32]. Through the use of the NLP pipeline, descriptions of symptom clusters, such as anhedonia, psychomotor retardation, and psychosocial stressors, can be more easily extracted, offering a more comprehensive understanding of patient experiences. This approach addresses the challenges inherent in psychiatric data, where nonstandardized language can obscure critical details about each patient's depression severity or clinical features. By extracting and categorizing these domains via NLP, researchers gain a more nuanced view of how depression manifests and evolves in real-world settings, supporting calls for thorough documentation in clinical practice.

**Figure 4.** Schematic representing a network of the "nine pillars of depression". Each node represents a key symptom or domain—such as anhedonia, depressed mood, weight changes, and sleep disturbances—which collectively define depression as a clinical syndrome. The network schematic emphasizes how these interconnected elements comprise the overarching concept of depression [26].




The depression symptom cluster includes the following symptom domains:

1. Anhedonia
2. Depressed mood
3. Changes in weight or appetite
4. Sleep disturbances
5. Psychomotor agitation or retardation
6. Feelings of anxiety
7. Feelings of guilt or worthlessness
8. Fatigue or diminished concentration
9. Suicidal ideation or thoughts of self-harm

## Challenges in Real-World Psychiatric Research

Despite its advantages, leveraging RWE in psychiatry presents notable challenges. A primary issue is the inconsistent documentation of critical clinical parameters [6]. Depression severity, for example, is often recorded in free-text notes without the use of standardized scales such as the Montgomery-Åsberg Depression Rating Scale (MADRS) or the Clinical Global Impression - Severity (CGI-S) scale [8]. This lack of standardization complicates the extraction and analysis of severity data [9,10]. With minimum extra work on behalf of the physicians, that is, noting down CGI-S scores, the clinical added value to be gained from patient records would be notable.

Unstructured clinical notes, while rich in contextual information, pose additional difficulties for data extraction. NLP tools must contend with variability in language, medical jargon, and incomplete or ambiguous entries. Recent advancements in NLP have improved the accuracy of concept extraction, yet challenges remain, particularly in capturing nuanced information such as reasons for treatment changes or patient-reported outcomes [6,33].

The integration of structured and unstructured data also requires careful attention to data quality and consistency. Discrepancies between these sources can lead to errors in analysis or misinterpretation of findings. For example, structured data may indicate a medication switch, but unstructured notes may reveal that the switch was due to adverse effects rather than lack of efficacy [6,8,27].

Our findings also revealed significant variability in the documentation and conceptualization of TRD [3]. We defined TRD as the initiation of a third antidepressant during the study period, as identified through structured data sources. However, unstructured notes frequently included additional nuances, such as mentions of partial response, adverse effects, or patient-reported discontinuation reasons. While the EMA's definition of TRD relies on two failed antidepressants in one same depressive episode [3], medication failure (eg, due to side effects or inefficacy) was not reliably captured in structured data alone. NLP contributed by identifying unstructured mentions of TRD that provided context for treatment changes, supporting the ongoing refinement of TRD criteria for clinical use.

## Limitations of the MOOD Project

While the MOOD project represents a significant step forward, several limitations must be acknowledged. First, the study relied on data from two hospitals in Belgium, so future expansions to include additional sites and more diverse datasets could broaden the extent of the findings. Second, the accuracy of the NLP pipeline, while robust, is contingent on the quality and completeness of the underlying data. Missing or ambiguous information in clinical notes can affect the reliability of extracted concepts, and variability in clinician terminology limits the accuracy of NLP tools, a challenge noted across studies



XSL•FO
RenderX

leveraging NLP for mental health interventions [6]. For instance, this analysis did not attempt to algorithmically infer or predict depression severity scores when not documented; severity capture was descriptive and limited to explicitly recorded severity descriptors and scale measurements identified in structured fields or narrative notes. Efforts to standardize documentation practices and improve data curation are critical for addressing these issues [9]. Finally, the lack of patient-reported outcomes in the study highlights a gap in understanding the subjective experiences of patients with MDD and TRD. Incorporating patient-reported measures in future studies could provide a more comprehensive view of treatment efficacy and quality of life [33].

## Conclusions

The findings from the MOOD project underscore the importance of continued innovation in RWE methodologies for psychiatry. Expanding the study to additional hospitals and incorporating standardized documentation practices will enhance the generalizability of the results. RWE studies also have the potential to improve clinical decision-making. By uncovering patterns in real-world treatment use, such as the adoption of esketamine for TRD, these studies can guide the development of more effective and personalized care strategies [8,34,35]. Moreover, the ability to analyze patient trajectories in diverse populations can help address disparities in access to and outcomes of psychiatric care [9,34]. Furthermore, integrating patient-reported outcomes and exploring the use of novel data sources, such as wearable devices or mobile health applications, could enrich future research [33,36]. Collaboration between clinicians, data scientists, and policymakers is essential to ensure that RWE studies address the most pressing challenges in psychiatric care. By leveraging advancements in artificial intelligence and NLP, future projects can build on the foundation laid by the MOOD project, ultimately improving outcomes for patients with MDD, TRD, and other mental health conditions. Therefore, the MOOD project can be a first cornerstone for cross-border collaboration on a structured RWE network for mental health.

## Acknowledgments

We thank Prof Giovanni Briganti for his thorough revision and editing of the manuscript, as well as Clara L. Oeste for medical writing.

## Funding

This study was funded by Johnson & Johnson.

## Data Availability

The datasets generated or analyzed during this study are not publicly available due to them being property of each participating hospital and therefore subject to strict privacy laws, but are available from the corresponding author on reasonable request.

## Conflicts of Interest

LynxCare reports that financial support was provided by Johnson & Johnson. DZ, KB, GDB, and IC report that financial support was provided by LynxCare. FVD is an employee at LynxCare. CE was an employee at LynxCare. DS and EP are employees at Johnson & Johnson.

Multimedia Appendix 1
Additional tables.
[DOCX File , 74 KB - ijmr_v15i1e86448_app1.docx ]

## References

1. Friedrich MJ. Depression is the leading cause of disability around the world. JAMA 2017;317(15):1517. [doi: 10.1001/jama.2017.3826] [Medline: 28418490]
2. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 2006;163(11):1905-1917. [doi: 10.1176/ajp.2006.163.11.1905] [Medline: 17074942]
3. McIntyre R, Alsuwaidan M, Baune B, Berk M, Demyttenaere K, Goldberg J, et al. Treatment-resistant depression: definition, prevalence, detection, management, and investigational interventions. World Psychiatry 2023;22(3):394-412 [FREE Full text] [doi: 10.1002/wps.21120] [Medline: 37713549]
4. Perlis RH, Iosifescu DV, Castro VM, Murphy SN, Gainer VS, Minnier J, et al. Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. Psychol Med 2012;42(1):41-50 [FREE Full text] [doi: 10.1017/S0033291711000997] [Medline: 21682950]
5. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence - what is it and what can it tell us? N Engl J Med 2016;375(23):2293-2297. [doi: 10.1056/NEJMsb1609216] [Medline: 27959688]



XSL•FO
RenderX

6. Malgaroli M, Hull TD, Zech JM, Althoff T. Natural language processing for mental health interventions: a systematic review and research framework. Transl Psychiatry 2023;13(1):309 [FREE Full text] [doi: 10.1038/s41398-023-02592-2] [Medline: 37798296]
7. Vaci N, Liu Q, Kormilitzin A, De Crescenzo F, Kurtulmus A, Harvey J, et al. Natural language processing for structuring clinical text data on depression using UK-CRIS. Evid Based Ment Health 2020;23(1):21-26 [FREE Full text] [doi: 10.1136/ebmental-2019-300134] [Medline: 32046989]
8. Wang H, Lyu N, Huang J, Fu B, Shang L, Yang F, et al. Real-world evidence from a retrospective study on suicide during depression: clinical characteristics, treatment patterns and disease burden. BMC Psychiatry 2024;24(1):300 [FREE Full text] [doi: 10.1186/s12888-024-05726-y] [Medline: 38641767]
9. Kern DM, Cepeda MS, Defalco F, Etropolski M. Treatment patterns and sequences of pharmacotherapy for patients diagnosed with depression in the United States: 2014 through 2019. BMC Psychiatry 2020;20(1):4 [FREE Full text] [doi: 10.1186/s12888-019-2418-7] [Medline: 31900133]
10. Cepeda MS, Reps J, Ryan P. Finding factors that predict treatment-resistant depression: results of a cohort study. Depress Anxiety 2018;35(7):668-673 [FREE Full text] [doi: 10.1002/da.22774] [Medline: 29786922]
11. Anderson HD, Kim K, Zhang M, Gutierrez E, Malhotra J, Tak C. Assessing availability of depression severity indicators in electronic health record data: a retrospective study in two large academic health care systems in the United States. Psychol Serv 2026;23(1):111-120. [doi: 10.1037/ser0000969] [Medline: 40455496]
12. Hens D, Wyers L, Claeys KG. Validation of an artificial intelligence driven framework to automatically detect red flag symptoms in screening for rare diseases in electronic health records: hereditary transthyretin amyloidosis polyneuropathy as a key example. J Peripher Nerv Syst 2023;28(1):79-85. [doi: 10.1111/jns.12523] [Medline: 36468607]
13. Moya A, Oeste CL, Beles M, Verstreken S, Dierckx R, Heggermont W, et al. Detection of transthyretin amyloid cardiomyopathy by automated data extraction from electronic health records. ESC Heart Fail 2023;10(6):3483-3492 [FREE Full text] [doi: 10.1002/ehf2.14517] [Medline: 37726928]
14. Beles M, Masuy I, Verstreken S, Bartunek J, Dierckx R, Heggermont W, et al. Cardio-renal-metabolic syndrome: clinical features and dapagliflozin eligibility in a real-world heart failure cohort. ESC Heart Fail 2023;10(4):2269-2280 [FREE Full text] [doi: 10.1002/ehf2.14381] [Medline: 37095712]
15. Van de Meulebroucke C, Beckers J, Corten K. What can we expect following anterior total hip arthroplasty on a regular operating table? a validation study of an artificial intelligence algorithm to monitor adverse events in a high-volume, nonacademic setting. J Arthroplasty 2019;34(10):2260-2266. [doi: 10.1016/j.arth.2019.07.039] [Medline: 31445868]
16. Crema C, Attardi G, Sartiano D, Redolfi A. Natural language processing in clinical neuroscience and psychiatry: a review. Front Psychiatry 2022;13:946387 [FREE Full text] [doi: 10.3389/fpsyt.2022.946387] [Medline: 36186874]
17. Guideline on clinical investigation of medicinal products in the treatment of depression. European Medicines Agency. 2023. URL: https://tinyurl.com/bdhb9b6y [accessed 2025-06-23]
18. Spravato (esketamine). European Medicines Agency. 2019. URL: https://www.ema.europa.eu/en/documents/overview/spravato-epar-medicine-overview_en.pdf [accessed 2026-01-26]
19. Gaynes BN, Warden D, Trivedi MH, Wisniewski SR, Fava M, Rush AJ. What did STAR*D teach us? results from a large-scale, practical, clinical trial for patients with depression. Psychiatr Serv 2009;60(11):1439-1445. [doi: 10.1176/ps.2009.60.11.1439] [Medline: 19880458]
20. Ionescu DF, Rosenbaum JF, Alpert JE. Pharmacological approaches to the challenge of treatment-resistant depression. Dialogues Clin Neurosci 2022;17(2):111-126. [doi: 10.31887/dcns.2015.17.2/dionescu]
21. Brown S, Rittenbach K, Cheung S, McKean G, MacMaster FP, Clement F. Current and common definitions of treatment-resistant depression: findings from a systematic review and qualitative interviews. Can J Psychiatry 2019;64(6):380-387 [FREE Full text] [doi: 10.1177/0706743719828965] [Medline: 30763119]
22. Gill K, Hett D, Carlish M, Amos R, Khatibi A, Morales-Muñoz I, et al. Examining the needs, outcomes and current treatment pathways of 2461 people with treatment-resistant depression: mixed-methods study. Br J Psychiatry 2026;228(2):108-115. [doi: 10.1192/bjp.2024.275] [Medline: 40070346]
23. The Book of OHDSI: Observational Health Data Sciences and Informatics. New York, NY: OHDSI; 2019.
24. Delombaerde D, Oeste CL, Geldhof V, Croes L, Bassez I, Verbiest A, et al. Cardiovascular toxicities in cancer patients treated with immune checkpoint inhibitors: multicenter study using natural language processing on Belgian hospital data. ESMO Real World Data Digit Oncol 2025;7:100111 [FREE Full text] [doi: 10.1016/j.esmorw.2024.100111] [Medline: 41647338]
25. Vanderkerken M, Van Eygen K, Galle V, Verbiest A, Janssens A, Masuy I, et al. Leveraging digital technology and artificial intelligence to describe the real-world belgian chronic lymphocytic leukemia patient population: the BE-CLLEAR Study. JCO Clin Cancer Inform 2026;10:e2500159. [doi: 10.1200/CCI-25-00159] [Medline: 41849725]
26. Kramchaninova A, Oeste C, Farokhshad N, Sterckx L. Biomedical LLMs pretrained on non-EHR data underperform in multilingual real-world settings. 2025 Presented at: Proceedings of the Workshop on Medical Language Processing in the Era of Large Language Models (MLP-LLM 2025); June 30-July 4, 2025; Marseille, France p. 25-41 URL: https://aclanthology.org/2025.jeptalnrecital-mlpllm.3.pdf




XSL•FO
RenderX

27. McDaniel BT, Cornet V, Carroll J, Chrones L, Chudzik J, Cochran J, et al. Real-world clinical outcomes and treatment patterns in patients with MDD treated with vortioxetine: a retrospective study. BMC Psychiatry 2023;23(1):938 [FREE Full text] [doi: 10.1186/s12888-023-05439-8] [Medline: 38093196]
28. Heerlein K, Young AH, Otte C, Frodl T, Degraeve G, Hagedoorn W, et al. Real-world evidence from a European cohort study of patients with treatment resistant depression: baseline patient characteristics. J Affect Disord 2021;283:115-122 [FREE Full text] [doi: 10.1016/j.jad.2020.11.124] [Medline: 33545659]
29. Heerlein K, Perugi G, Otte C, Frodl T, Degraeve G, Hagedoorn W, et al. Real-world evidence from a European cohort study of patients with treatment resistant depression: treatment patterns and clinical outcomes. J Affect Disord 2021;290:334-344 [FREE Full text] [doi: 10.1016/j.jad.2021.03.073] [Medline: 34044256]
30. Gillain B, Degraeve G, Dreesen T, De Bruecker G, Buntinx E, Beke D, et al. Real-world treatment patterns, outcomes, resource utilization and costs in treatment-resistant major depressive disorder: pattern, a retrospective cohort study in Belgium. Pharmacoecon Open 2022;6(2):293-302 [FREE Full text] [doi: 10.1007/s41669-021-00306-2] [Medline: 34782984]
31. Diagnostic and Statistical Manual of Mental Disorders. Washington DC: American Psychiatric Association; 2013.
32. Fried E, Nesse R. Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR*D study. J Affect Disord 2015;172:96-102 [FREE Full text] [doi: 10.1016/j.jad.2014.10.010] [Medline: 25451401]
33. Maruszczyk K, Aiyegbusi OL, Cardoso VR, Gkoutos GV, Slater K, Collis P, et al. Implementation of patient-reported outcome measures in real-world evidence studies: analysis of clinicalTrials.gov records (1999-2021). Contemp Clin Trials 2022;120:106882 [FREE Full text] [doi: 10.1016/j.cct.2022.106882] [Medline: 35973663]
34. Cepeda MS, Kern DM, Canuso CM. At baseline patients treated with esketamine have higher burden of disease than other patients with treatment resistant depression: learnings from a population based study. Depress Anxiety 2021;38(5):521-527 [FREE Full text] [doi: 10.1002/da.23138] [Medline: 33475213]
35. Ochs-Ross R, Daly EJ, Zhang Y, Lane R, Lim P, Morrison RL, et al. Efficacy and safety of esketamine nasal spray plus an oral antidepressant in elderly patients with treatment-resistant depression-TRANSFORM-3. Am J Geriatr Psychiatry 2020;28(2):121-141 [FREE Full text] [doi: 10.1016/j.jagp.2019.10.008] [Medline: 31734084]
36. Appleton R, Williams J, Vera San Juan N, Needle JJ, Schlief M, Jordan H, et al. Implementation, adoption, and perceptions of telemental health during the covid-19 pandemic: systematic review. J Med Internet Res 2021;23(12):e31746 [FREE Full text] [doi: 10.2196/31746] [Medline: 34709179]


## Abbreviations

**CDM:** common data model
**CGI-S:** clinical global impressions-severity
**CUI:** concept unique identifier
**DSM-IV:** Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition)
**DSM-V:** Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
**EHR:** electronic health record
**EMA:** European Medicines Agency
**ICD:** International Classification of Diseases
**ICD-10:** International Statistical Classification of Diseases, Tenth Revision
**MADRS:** Montgomery-Åsberg Depression Rating Scale
**MDD:** major depressive disorder
**NLP:** natural language processing
**OMOP:** Observational Medical Outcomes Partnership
**NER:** named entity recognition
**OOTB:** out-of-the-box
**RWE:** real-world evidence
**SNOMED CT:** Systematized Nomenclature of Medicine – Clinical Terms
**TRD:** treatment-resistant depression
**UMLS:** Unified Medical Language System



XSL•FO
RenderX

*Edited by M Balcarras; submitted 24.Oct.2025; peer-reviewed by D Chrimes, U Pasupulety; comments to author 07.Jan.2026; revised version received 19.Mar.2026; accepted 19.Mar.2026; published 22.May.2026.*

*Please cite as:*
*Zeeuws D, Bernagie K, De Bruecker G, Claeys I, Evenepoel C, Ver Donck F, Smeets D, Peeters E*
*Real-World Clinical Characterization of Major Depressive Disorder and Treatment-Resistant Depression Supported by Natural Language Processing: Multicenter Observational Study From the MOOD Project*
*Interact J Med Res 2026;15:e86448*
*URL: https://www.i-jmr.org/2026/1/e86448*
*doi:10.2196/86448*
*PMID:*

©Dieter Zeeuws, Katrien Bernagie, Geert De Bruecker, Isabel Claeys, Charlotte Evenepoel, Fabienne Ver Donck, David Smeets, Elke Peeters. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 22.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Digital Interventions Addressing Cognitive and Psychological Symptoms in Long COVID: Scoping Review of Multicomponent Approaches

Sandra León-Herrera[1], PhD; Marta Sánchez-Castro[2], PhD; Ana Luisa Neves[3,4]; Mª Pilar Rodríguez-Pérez[5], PhD; Vinicius Jobim Fischer[6], PhD; Djenna Hutmacher[6]; Reham Aldakhil[3]; Marina Vaillancourt de Dios[7]; Vinicius Anjos de Almeida[8]; Bárbara Oliván-Blázquez[1], PhD; Rosa Magallón-Botaya[9], PhD; Charles Benoy[6,10]; Raquel Gómez-Bravo[6,11], PhD

[1]Department of Psychology and Sociology, University of Zaragoza, Zaragoza, Spain

[10]Universitäre Psychiatrische Kliniken Basel, Basel, Switzerland

[11]Research Group Self-Regulation and Health, Institute for Health and Behaviour, Department of Behavioural and Cognitive Sciences, Faculty of Humanities, Education, and Social Sciences, Luxembourg University, Luxembourg, Luxembourg

[2]Department of Life Sciences and Medicine, University of Luxembourg, Luxembourg, Luxembourg

[3]Department of Primary Care and Public Health, Imperial College London, South Kensington Campus London, London, United Kingdom

[4]RISE-Health, University of Porto, Porto, Portugal

[5]Department of Physiotherapy, Occupational Therapy, Rehabilitation and Physical Medicine, University Rey Juan Carlos, Madrid, Spain

[6]Rehaklinik, Centre Hospitalier Neuro-psychiatrique (CHNP), Ettelbruck, Luxembourg

[7]Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, The Netherlands

[8]University of Säo Paulo, Säo Paulo, Brazil

[9]Department of Medicine, Faculty of Medicine, University of Zaragoza, Zaragoza, Spain

**Corresponding Author:**
Ana Luisa Neves
Department of Primary Care and Public Health, Imperial College London, South Kensington Campus London, London, United Kingdom

## Abstract

**Background:** Long COVID, or postacute COVID-19 syndrome, presents with persistent cognitive and psychological symptoms such as *brain fog*, anxiety, depression, and fatigue, significantly impacting quality of life and daily functioning. Digital health interventions offer a scalable, accessible solution to bridge care gaps, especially where conventional neuropsychological support is limited. However, evidence regarding their effectiveness for neuropsychiatric symptoms in long COVID remains fragmented.

**Objective:** This scoping review aimed to systematically identify and map the existing evidence on digital interventions targeting cognitive and psychological symptoms in individuals with long COVID. The review also sought to categorize intervention types, assess reported outcomes, and identify methodological gaps to inform future clinical and research priorities.

**Methods:** The review followed the Arksey and O'Malley framework and adhered to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines. Comprehensive searches were conducted in 4 databases (PubMed, Scopus, Web of Science, and ScienceDirect) from December 2024 to February 2025. Eligible studies included peer-reviewed and gray literature published in English or Spanish since 2020. Studies were screened and selected based on predefined inclusion and exclusion criteria. Data were extracted using a standardized charting form and synthesized narratively, with thematic grouping by intervention type.

**Results:** Of 888 records identified, 25 (2.82%) were included. Intervention types encompassed telehealth platforms, mobile health apps, virtual reality, online cognitive and psychological therapies, game-based cognitive training, neuromodulation (transcranial direct current stimulation), and multicomponent programs. Most studies reported improvements in psychological well-being, emotional regulation, and cognitive domains such as attention and memory. However, findings varied, with some interventions showing no significant cognitive gains or sustained effects. Common limitations included small sample sizes, lack of control groups, heterogeneity in outcomes and intervention protocols, and short follow-up durations. The underrepresentation of older adults and underserved populations was also noted.

**Conclusions:** Digital interventions show promise for addressing cognitive and psychological symptoms in long COVID, particularly when delivered as multicomponent programs. Nonetheless, the evidence base remains preliminary. Future research



XSL•FO
RenderX

should prioritize high-quality randomized trials with standardized outcome measures, long-term follow-up, and diverse participant samples. Addressing barriers related to digital literacy and access will be essential to ensure equity and real-world effectiveness.

**Trial Registration:** OSF Registries 10.17605/OSF.IO/HX7UE; https://osf.io/hx7ue/overview

***(Interact J Med Res 2026;15:e80616)*** doi:10.2196/80616

**KEYWORDS**

long COVID; postacute COVID-19 syndrome; digital health interventions; cognitive symptoms; psychological symptoms; telehealth; eHealth

## Introduction

The COVID-19 pandemic has resulted in an unprecedented global health crisis, with more than 700 million confirmed cases worldwide to date. Beyond the acute phase [1,2], a substantial number of individuals experience persistent symptoms, collectively referred to as *long COVID* or *postacute COVID-19 syndrome* [3,4], which can persist for weeks or months after initial recovery. Among the most reported and debilitating effects are cognitive and psychological impairments [5,6], which significantly impair daily functioning, employment, and overall quality of life.

Cognitive symptoms, colloquially referred to as *brain fog*, include problems with attention, memory, executive functioning, and information processing [7,8]. These manifestations can severely impair daily functioning, work performance, and social participation [9,10]. Concurrently, psychological symptoms such as anxiety, depression, posttraumatic stress, sleep disturbances, and fatigue are highly prevalent in individuals with long COVID, further compounding the burden of disease. These neuropsychiatric sequelae often persist long after the resolution of respiratory symptoms and may occur even in individuals who experienced only mild acute illness [11,12].

Emerging evidence suggests that the mechanisms underlying these symptoms are complex and multifactorial, including neuroinflammation, microvascular damage, immune dysregulation, and psychosocial stressors [13,14]. The unpredictability of symptom progression and the lack of clear diagnostic markers pose challenges for clinical management and rehabilitation planning. Furthermore, many patients report difficulties in accessing appropriate care, particularly neuropsychological and mental health services, due to geographical, logistical, or systemic barriers [15,16].

In this context, digital health interventions have gained considerable attention as innovative and scalable tools to bridge gaps in care. These interventions encompass a wide range of modalities, including mobile health (mHealth) apps, telehealth consultations, online therapy platforms, virtual reality (VR) tools, and other web-based cognitive or behavioral training programs [17]. Digital interventions offer the potential for continuous, remote, and personalized support, which is particularly valuable for populations affected by chronic illness and mobility limitations [18].

Although there is a growing body of research exploring the use of digital health tools in the context of long COVID, most studies to date have addressed general management or physical recovery. Fewer have specifically focused on interventions targeting the cognitive and psychological symptoms that are among the most disabling and persistent sequelae of the condition. As such, health care professionals lack a synthesized and focused body of evidence to inform clinical decision-making in this domain, and researchers face uncertainty regarding which digital approaches show the most promise for addressing neuropsychological aspects of long COVID [17,19,20]. Given the rapid evolution of digital health technologies and the fragmented nature of the current literature, a comprehensive mapping of existing evidence in this area is urgently needed [21]. This is particularly important given the increasing prevalence of long COVID and the growing demand for accessible, nonpharmacological support strategies.

This scoping review aims to address this knowledge gap by systematically identifying, mapping, and synthesizing the existing evidence on digital interventions targeting cognitive and psychological symptoms in individuals with long COVID, while categorizing intervention types, evaluating reported outcomes, and identifying key methodological gaps to inform future research and clinical practice.

## Methods

### Study Design

We conducted a scoping review to examine the evidence on digital interventions addressing cognitive and psychological symptoms in individuals with long COVID. This methodology was chosen due to the exploratory nature of the research question and the heterogeneity of the available evidence in terms of study designs, intervention types, and outcome measures. Scoping reviews are particularly suited to emerging research areas, as they allow a comprehensive overview of the literature and the identification of key knowledge gaps, rather than a narrowly focused assessment of effectiveness.

This review followed the framework proposed by Arksey and O'Malley [22], which comprises 5 key stages: identifying the research question; identifying relevant studies; selecting studies; charting the data; and collating, summarizing, and reporting the results. To enhance transparency and methodological rigor, the review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines [23]. The completed PRISMA-ScR checklist is provided as Checklist 1. Additionally, a publicly accessible study protocol was registered in the Open Science Framework on December 11, 2024 [24].



XSL•FO
RenderX

## Research Questions

To structure the research question, we used the population, intervention, comparison, outcomes, and study design framework [25], summarized in Textbox 1.

**Textbox 1.** Population, intervention, comparison, outcomes, and study design (PICOS) criteria for the research question.

**Population**

- Individuals with long COVID who experience cognitive or psychological symptoms

**Intervention**

- Digital interventions (eg, mobile apps, virtual reality, telehealth, and online cognitive or psychological therapies)

**Comparison**

- No comparison was required for inclusion; comparators, when present, were reported

**Outcomes**

- Cognitive outcomes (eg, attention, memory, and executive function) and psychological outcomes (eg, depression, anxiety, fatigue, and well-being)

**Study design**

- Randomized controlled trials, observational studies, protocols, qualitative studies, and reviews (including scoping, systematic, and narrative reviews), as well as gray literature

The secondary research questions were as follows:

- What digital interventions are available and effective for managing cognitive and psychological symptoms in individuals with long COVID?
- What types of digital interventions are most commonly used?
- What evidence supports their efficacy?
- What methodological limitations and knowledge gaps exist?

## Identifying Relevant Studies

A comprehensive search strategy was designed and peer-reviewed by a research librarian. The following databases were searched between December 2024 and February 2025: PubMed, Scopus, Web of Science, and ScienceDirect. Search terms combined concepts related to long COVID (eg, “Long COVID,” “post-COVID,” “post-acute COVID-19 syndrome,” and “chronic COVID”), digital interventions (eg, “teletherapy,” “digital intervention,” “mobile app,” “telehealth,” “eHealth,” and “virtual reality”), and neuropsychiatric symptoms (eg, “cognitive symptoms,” “*brain fog*,” “anxiety,” “depression,” and “mental health”).

In addition to peer-reviewed records, our search strategy included gray literature available within the databases consulted—particularly ScienceDirect—which indexes materials such as book chapters, conference proceedings, and scientific reports. These records were screened using the same inclusion and exclusion criteria applied to peer-reviewed studies.

The search was limited to studies published in English or Spanish from 2020 onward to capture literature emerging during the COVID-19 pandemic. These languages were selected based on the language competencies of the review team and the high likelihood of relevant literature being published in these languages during the COVID-19 pandemic.

The complete search strings for each database are available in Multimedia Appendix 1 [26,27].

Although hand-searching of reference lists is commonly used as an additional step to identify relevant studies, it was not performed in this review. This decision was based on the comprehensiveness of the database search strategy, which was developed in collaboration with a research librarian and included multiple major databases and gray literature sources. This approach was considered sufficient to capture the relevant body of evidence.

## Study Selection

Study selection was conducted in 2 stages. Initially, 2 reviewers independently assessed titles and abstracts to evaluate their potential relevance. Studies deemed potentially eligible were then examined in full text. Disagreements between reviewers were addressed through discussion based on predefined inclusion criteria (Textbox 2), and a third reviewer was involved when necessary to achieve consensus.

No restrictions were placed on study design or sample size. All identified relevant studies were included to provide a comprehensive overview.



XSL•FO
RenderX

**Textbox 2.** Eligibility criteria for the study.

**Inclusion criteria**

- Original studies (quantitative, qualitative, and mixed methods), reviews (including systematic, scoping, and narrative reviews), or study protocols
- Studies involving digital interventions targeting cognitive and/or psychological symptoms in patients with long COVID, either exclusively or as part of a broader multicomponent approach
- Studies published from 2020 onward in English or Spanish

**Exclusion criteria**

- Studies focusing solely on physical symptoms or nondigital interventions
- Studies not reporting outcomes related to cognitive or psychological health
- Studies published in languages other than English or Spanish

### Data Charting

Data from the selected studies were charted using a standardized data extraction table designed to capture key information related to the research questions. The extracted data included authors, year of publication, study design and setting, population and sample size, type and description of the digital intervention (including duration and frequency), cognitive and/or psychological outcomes assessed, key findings, and reported limitations. The data extraction form was piloted on a subset of 3 studies and refined to improve consistency and capture all relevant variables.

Data were extracted by one reviewer and verified by a second reviewer. Discrepancies were resolved by consensus. The charting process remained iterative, allowing for the inclusion of additional variables as needed.

### Data Synthesis

A narrative synthesis was performed to describe the digital interventions, their outcomes, and methodological characteristics. A descriptive thematic analysis was used to classify interventions into key categories (eg, VR, teletherapy, and mobile apps). Outcomes were synthesized narratively according to their effects on cognitive and psychological domains. The interventions were grouped thematically and compared based on their reported effects on cognitive and psychological symptoms.

Findings are presented in tabular format and complemented by descriptive summaries. Additionally, key gaps in the literature and recommendations for future research were outlined based on patterns identified during synthesis.

Because this is a scoping review, no formal quality assessment or meta-analysis was conducted, in line with best practices for this review type. However, reported limitations and methodological concerns were documented.

### Ethical Considerations

This study involved the review of publicly available literature and did not require ethics approval. No human participants or personal data were involved. This study complies with ethical standards for secondary research and adheres to the principles outlined in the Declaration of Helsinki, where applicable.

## Results

### Sources Selection and Characteristics

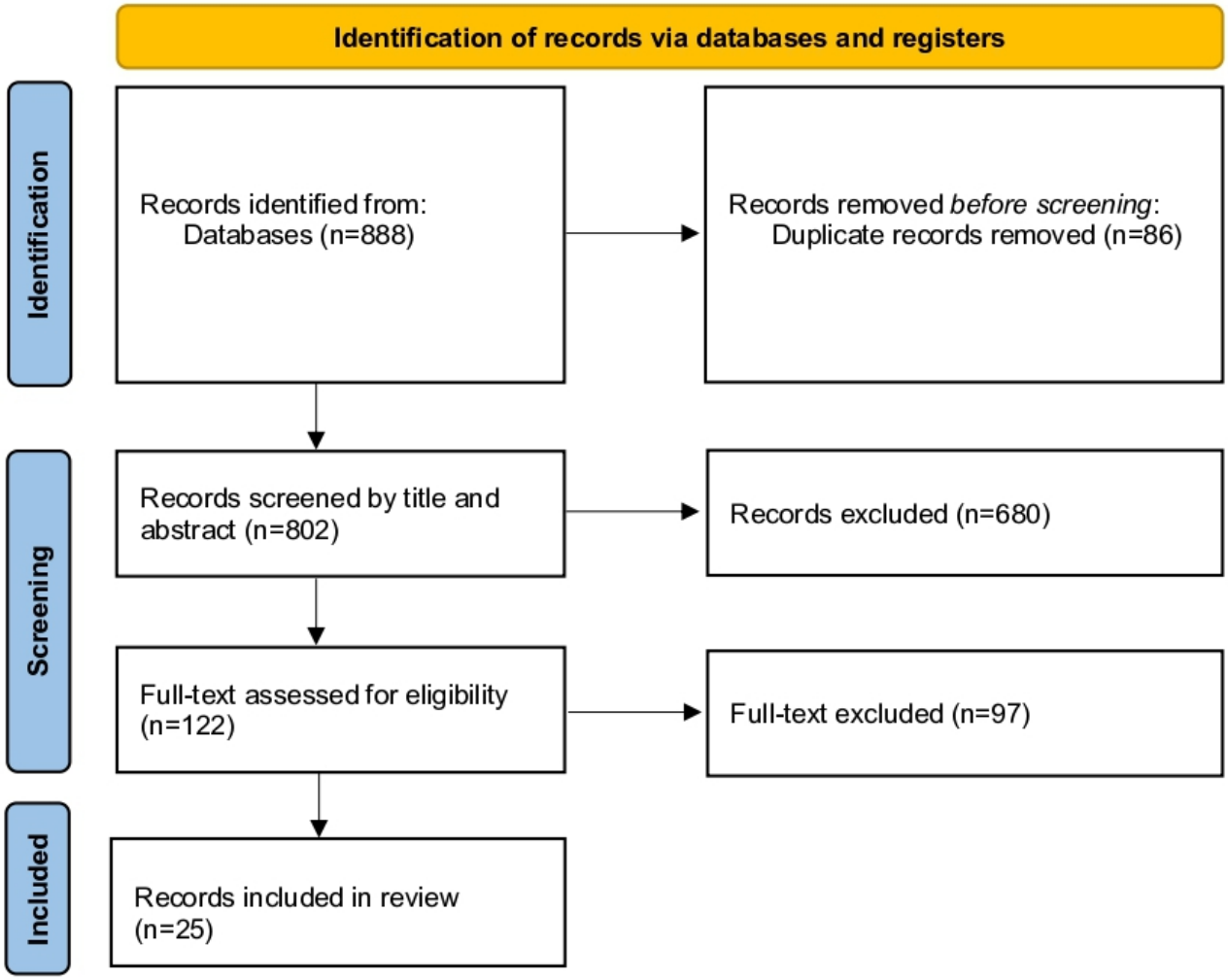
The initial database searches yielded 888 records. After removing duplicates, 802 (90.32%) unique records remained for title and abstract screening. Of these, 680 (76.58%) records were excluded based on predefined inclusion and exclusion criteria. In total, 122 (13.74%) full-text records were retrieved and assessed for eligibility, leading to the exclusion of 97 (10.92%) records for reasons such as irrelevance to cognitive or psychological symptoms, lack of a digital intervention, or publication in an unsupported language. Following full-text screening, 25 (2.82%) records were included in this scoping review. These records included peer-reviewed research papers, book chapters, and other relevant literature. Although gray literature (eg, commentaries, editorials, and policy statements) was considered, only the abovementioned records met the inclusion criteria.

The selection process is illustrated in the PRISMA-ScR flow diagram (Figure 1), summarizing the records identified, screened, included, and excluded. In total, 8 (32%) studies were study protocols and had not yet reported outcome data at the time of the review.



XSL•FO
RenderX

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.




To improve readability and facilitate comparison across records, a simplified summary of the included interventions is provided in Table 1. This table groups the records by type of digital intervention, lists the corresponding authors, and highlights the main reported effects on cognitive and psychological symptoms. Detailed methodological characteristics and outcome measures for each study can be found in Multimedia Appendix 2.



XSL•FO
RenderX

**Table .** Summary of digital interventions and main reported effects.

| Records (author, reference, and year) | Intervention type | Main reported effects |
|---|---|---|
| León-Herrera et al [28], 2024; Lai et al [29], 2024; Hatcher et al [30], 2022; Müllenmeister et al [31], 2024; Smith et al [32], 2023; Al-Jabr et al [33], 2022; Ennis et al [34], 2023; and Vanova et al [35], 2024 | Telehealth or telerehabilitation | Improvements in quality of life, mental health (anxiety and depression), and self-efficacy, and mixed effects on cognition |
| Widmann et al [36], 2024; Cano et al [37], 2024; Stavrou et al [38], 2023; Groenveld et al [39], 2022; Prudent et al [40], 2024; and Krotz et al [41], 2023 | Virtual reality | Enhanced emotional well-being and physical functioning, inconsistent effects on cognition, and some usability issues |
| Daniels et al [19], 2024; Mata et al [42], 2023; Smith et al [32], 2023; and Montes-Ibarra et al [43], 2024 | Mobile health apps | Gains in attention, mood, and self-reported recovery, often combined with mindfulness or educational components |
| Tsang and Tabio [44], 2024; Prudent et al [40], 2024; and Al-Jabr et al [33], 2022 | Online psychological therapies | Decreased anxiety, posttraumatic stress disorder, and depression and high feasibility and acceptability |
| Victoria et al [45], 2024 | Game-based cognitive training | Improved processing speed and task-switching and no sustained improvement in attention |
| Eilam-Stock et al [46], 2021 | Neuromodulation (transcranial direct current stimulation) | Enhanced cognition, mood, and energy levels in a small-scale pilot study |
| León-Herrera et al [28], 2024; Smith et al [32], 2023; Daniels et al [19], 2024; Krotz et al [41], 2023; Schröder et al [47], 2023; and Dahmen et al [48], 2022 | Multicomponent programs | Broad improvements across mental health, cognitive symptoms, and physical activity |

## Population Characteristics

The studies included in this scoping review investigated individuals with persistent post–COVID-19 symptoms across various countries, including Germany, Spain, the United Kingdom, the United States, Canada, Greece, Italy, the Netherlands, and Taiwan. Most included studies were conducted in Europe and North America, indicating a geographic concentration of research and limited evidence from low- and middle-income countries. Populations commonly presented with a mix of cognitive symptoms (eg, *brain fog*, memory impairment, and concentration difficulties) and psychological symptoms (eg, anxiety, depression, fatigue, and sleep disturbances), along with impaired physical function and reduced quality of life.

Sample sizes varied significantly, from as few as 11 participants in a pilot telehealth group therapy study [44] to >600 participants in a blended digital rehabilitation program [32]. Some studies focused on specific age groups—for example, the CICERO trial [35] limited participants to adults aged 30 to 60 years—while others explored experiences across younger and older adults [42]. A common trend was the overrepresentation of women, with some studies noting that digital tool acceptance was higher among women and those with prior digital experience [47]. Although some studies reported participants' age, gender, and previous digital experience, comprehensive demographic data, such as digital literacy and socioeconomic background, were not systematically reported, limiting subgroup analysis and generalizability.

## Types of Digital Interventions

Digital health interventions varied widely in format, intensity, and content. The most common approaches included telehealth and online rehabilitation platforms, mHealth tools, VR, game-based cognitive therapies, digital psychotherapy, multicomponent programs, and neuromodulation techniques.

Telehealth and online rehabilitation platforms typically combined cognitive training, physical exercise, and mental health support delivered via weekly video calls and digital resources [28,29]. mHealth tools were also frequently implemented; for example, an immersive VR-based app integrating mindfulness and cognitive training was used [37], alongside mobile apps and wearable monitoring tools [19].

VR interventions were explored in several records, showing benefits in physical recovery and emotional well-being [36,38,39], although some users reported side effects such as dizziness and headaches. Game-based cognitive therapy was also represented by a video game designed to enhance cognitive function, which led to improvements in processing speed and executive function.

Digital psychotherapy approaches were also identified, including trauma-focused cognitive behavioral therapy (CBT) delivered online [40] and telehealth group CBT incorporating verified COVID-19 information and peer support [44]. In addition, multicomponent programs combined web-based exercise classes, symptom education, and in-person community support [32]. Finally, neuromodulation techniques, including remotely supervised transcranial direct current stimulation, were reported to improve fatigue and depression in a small case series [46].



XSL•FO
RenderX

The duration and frequency of interventions ranged from 2 weeks to 3 months, with varying intensity and adherence strategies, such as weekly sessions or daily app use.

### Intervention Efficacy

Outcomes evaluated across records spanned several domains. Cognitive function was frequently assessed, including memory, executive function, attention, and processing speed [35,37,45]. Psychological health outcomes were also commonly reported, particularly depression, anxiety, posttraumatic stress disorder, sleep quality, and fatigue [28,40].

In addition, several records examined physical health indicators such as breathlessness, exercise tolerance, sleep quality, and overall physical performance [29,38]. Broader measures of quality of life and functioning—including self-efficacy, daily activities, occupational performance, and health care use—were also assessed [31,32]. Finally, usability and acceptability outcomes, such as satisfaction, adherence, and technological barriers, were reported in a subset of sources [39,47].

Many records reported positive effects across multiple outcome domains. Cognitive improvements were observed in several studies, including enhanced processing speed and attention [45], global cognition [37], and memory performance [28]. However, these effects were not consistent across all studies, with some reporting no significant cognitive gains [38]. Variability in intervention type, duration, and assessment tools likely contributed to these mixed findings.

Psychological benefits included reductions in depression and anxiety [28,40], as well as improvements in emotional well-being. Physical outcomes also improved in several trials, including enhanced sleep quality [29], decreased breathlessness [38], and increased physical activity [32]. Additionally, quality of life and self-efficacy improved, particularly in studies integrating behavioral support and health education [19,28].

Nonetheless, effects were not universally sustained, and several studies reported limited or nonsignificant changes in certain domains, such as sustained attention or long-term mental health outcomes.

### Methodological Limitations and Knowledge Gaps

The methodological quality and design heterogeneity of the included studies present important considerations for interpreting the findings of this review. Although several randomized controlled trials [28,29,45] contributed robust evidence, a substantial proportion of the literature consisted of pilot studies, observational designs, or study protocols without reported outcomes [30,34,48]. This variability limits the overall strength of conclusions and underscores the early, exploratory nature of research in this domain.

Sample sizes were frequently small, particularly in feasibility or case studies [44,46], reducing statistical power and generalizability. Several studies lacked control groups or used nonrandomized designs, making it difficult to establish causal relationships between interventions and outcomes. Blinding—both of participants and researchers—was often absent or not reported, introducing a potential risk of bias.

Additionally, many studies relied heavily on self-reported outcome measures to assess cognitive and psychological symptoms. Although these tools provide valuable patient-centered data, they are subject to recall bias and may not accurately reflect clinical change. The use of diverse instruments across studies also complicates cross-study comparisons. Furthermore, follow-up periods were generally short, limiting insights into the sustainability of treatment effects over time. Overall, although early findings are promising, methodological constraints suggest that current evidence should be interpreted cautiously.

Despite growing interest in digital health interventions for long COVID, important gaps remain that hinder the development of standardized, scalable treatment approaches. One key limitation is the lack of long-term follow-up data. Most studies evaluated outcomes immediately after the intervention or within a few weeks, leaving unanswered questions about the durability of effects, relapse rates, and maintenance strategies.

Intervention protocols were highly heterogeneous in terms of content, delivery mode, duration, and intensity, making it difficult to identify the most effective components or to compare results across studies. The absence of standardized outcome measures further complicates synthesis and meta-analysis efforts. Moreover, although digital interventions are often assumed to be widely accessible, few studies addressed issues related to digital literacy, internet access, or usability challenges—factors particularly relevant for older adults and underserved populations.

Certain groups, including older adults, individuals with severe or complex post–COVID-19 symptoms, and those from lower socioeconomic backgrounds, were underrepresented in most samples. Cultural and gender-specific factors affecting the acceptability and efficacy of interventions were also largely unexplored. Finally, few records examined the mechanisms through which digital interventions exert their effects, leaving open important questions about how improvements in cognitive or emotional functioning are achieved and whether they translate into functional recovery.

Collectively, these gaps highlight the need for well-designed, adequately powered trials with longer follow-up, consistent outcome measures, and inclusive recruitment strategies. Future research should also explore how to personalize digital interventions and integrate them sustainably into health systems to meet the needs of diverse populations affected by long COVID.

## Discussion

### Principal Findings

This scoping review synthesized the current landscape of digital interventions targeting cognitive and psychological symptoms in individuals with long COVID. The findings reflect a field in active development, with diverse intervention types, from telehealth to VR, showing potential benefits across cognitive, emotional, and functional domains [6,49]. Although symptoms such as *brain fog*, memory deficits, fatigue, and anxiety were consistently addressed, intervention efficacy varied widely,



XSL•FO
RenderX

underscoring the need for more tailored, evidence-based approaches.

A notable strength of this emerging field is the international scope of research, encompassing studies conducted across Europe, North America, and Asia. Despite differences in study design and health care infrastructure, there was considerable consistency in the reported symptom burden among participants. Cognitive dysfunction and psychological distress, such as anxiety, depression, and fatigue, were prevalent across cohorts. These findings are consistent with prior research that has identified neurological and psychiatric effects as central and disabling features of long COVID [50,51].

Digital interventions were diverse in format, encompassing telehealth platforms, mobile apps, VR, digital cognitive therapies, and online behavioral support programs. Interventions often addressed multiple domains simultaneously, incorporating physical, psychological, and cognitive components. This multimodal approach aligns well with the multisystemic nature of long COVID and reflects a growing emphasis on integrative, patient-centered care [52]. For example, programs like those described by León-Herrera et al [28] and Smith et al [32] combined exercise, cognitive rehabilitation, and education, yielding improvements in quality of life and mental health outcomes.

Despite these promising findings, the effectiveness of digital interventions was not uniform. Although several studies reported improvements in processing speed, attention, emotional well-being, physical activity, and sleep quality, other outcomes—such as sustained attention and long-term psychological recovery—showed inconsistent or nonsignificant effects. These mixed results may reflect the variability in intervention type and intensity, as well as individual differences in symptomatology, baseline functioning, and engagement levels. Previous research has consistently highlighted the heterogeneous nature of long COVID symptoms, with some individuals experiencing more persistent and debilitating effects than others. This variability in symptom severity, along with differences in how patients engage with interventions, may explain the inconsistencies observed across studies. Moreover, the complexity of long COVID, involving multiple systems of the body, suggests that a one-size-fits-all approach to digital interventions may not be equally effective for everyone [53].

The methodological quality of the included records varied widely. Although this scoping review did not conduct a formal risk of bias assessment, a substantial number of included studies used nonrandomized or uncontrolled designs, limiting the strength of causal inferences and contributing to heterogeneity in findings. Although randomized controlled trials provided valuable insights (eg, Victoria et al [45], León-Herrera et al [28], and Lai et al [29]), many studies were limited by small sample sizes, lack of control groups, or reliance on self-reported measures. In particular, the short follow-up periods commonly used across studies limited the ability to assess the durability of observed improvements. Furthermore, the reliance on subjective outcome measures, often without objective neurocognitive or physiological assessments, raises concerns about measurement reliability and the potential for response bias.

In terms of equity and access, several important gaps remain. Many interventions assumed a basic level of digital literacy and reliable access to internet-enabled devices—assumptions that may not hold true for older adults, people with disabilities, or those in socioeconomically disadvantaged settings [54,55]. Notably, long COVID and digital access disparities often intersect with age, socioeconomic status, and geography, especially as digital interventions are intended to bridge, rather than widen, these gaps. Few studies systematically addressed these barriers, and older adults were often underrepresented despite being disproportionately affected by both COVID-19 and digital exclusion. Studies have shown that older adults, individuals from low-income households, and those in rural areas are significantly less likely to have consistent access to internet-enabled devices or the digital literacy needed to engage with these tools. Designing low-bandwidth applications, providing user training, and incorporating hybrid support models may help mitigate these disparities and expand reach [56,57].

Importantly, the mechanisms through which digital interventions exert their effects remain poorly understood. Whether improvements in cognitive or emotional domains result from specific therapeutic content, increased social engagement, improved physical fitness, or a combination of these factors is still unclear. Recent investigations have begun exploring these mechanisms, although definitive conclusions remain elusive [58,59]. Understanding these mechanisms will be essential for refining and targeting future interventions.

Overall, this review reveals a rapidly developing but still nascent field. To build a more robust evidence base, future research should prioritize high-quality randomized trials with adequate sample sizes, long-term follow-up, and standardized outcome measures [60,61]. Equally important is the inclusion of diverse populations and the evaluation of implementation factors such as usability, adherence, and cost-effectiveness. Addressing these gaps will be essential to integrate digital health solutions into routine care pathways and to ensure that they are accessible and effective for all individuals affected by long COVID.

### Strengths and Limitations

This scoping review offers a comprehensive and up-to-date synthesis of digital interventions aimed at addressing long COVID, encompassing a wide range of technological approaches and international contexts. A major strength is the breadth of included study designs, which enables the mapping of both experimental and exploratory initiatives, as well as emerging clinical protocols. By including various forms of digital care—such as telehealth, mobile apps, VR, and neuromodulation—this review captures the evolving landscape of post–COVID-19 rehabilitation and highlights the multidisciplinary nature of digital health innovation.

However, several limitations must be acknowledged. First, the heterogeneity of study designs, populations, and outcome measures limited the ability to perform direct comparisons or draw definitive conclusions about intervention efficacy. Second, as with many scoping reviews, the aim was to map available



XSL•FO
RenderX

evidence rather than evaluate quality in depth; thus, a formal risk of bias assessment was not conducted. Finally, the inclusion of ongoing trials and study protocols, while useful for identifying research trends, means that many findings are preliminary and not yet peer-reviewed.

### Clinical and Policy Implications

Physicians and health care systems may consider testing scalable, low-cost digital solutions such as online CBT, symptom tracking apps, or telerehabilitation platforms as complements to conventional care. To ensure successful implementation, it is essential to provide patients with support for uptake, address barriers to access, and align digital interventions with existing service delivery pathways. Integrating these tools can expand the reach of health care and maintain its continuity, especially in underserved populations.

### Future Directions

Future research should build on these early insights by prioritizing well-powered randomized controlled trials that include diverse, representative populations and longer follow-up periods to assess the sustainability of intervention effects. There is a clear need for standardized outcome measures that capture the multidimensional nature of long COVID, including validated tools for cognitive, emotional, and physical health, as well as functional status and quality of life. Research should also explore adaptive and personalized digital interventions that can respond to individual symptom profiles and levels of digital literacy.

In addition, future research should consider mixed methods approaches, as qualitative studies can provide insights into patient experiences and inform the design of well-powered randomized controlled trials. Conversely, trends and patterns observed in large trials can guide research exploring the lived experiences of underrepresented or marginalized populations, including older adults, individuals living in rural or remote areas, and people with disabilities.

Moreover, implementation science approaches are needed to evaluate how digital interventions can be integrated into real-world health care systems, addressing barriers such as technological access, health care professional training, and user engagement. Cost-effectiveness analyses and health equity assessments will be crucial to ensure that digital solutions do not widen existing disparities. Finally, investigating the mechanisms of action—such as whether improvements stem from behavioral change, neuroplasticity, or social support—can help refine digital treatment strategies and optimize outcomes for individuals living with long COVID.

## Conclusions

This scoping review highlights the growing use of digital interventions—such as telehealth, mobile apps, VR, and online therapies—for managing cognitive and psychological symptoms in individuals with long COVID. Although early evidence suggests potential benefits across domains such as mood, attention, and quality of life, findings remain heterogeneous, and sustained improvements are inconsistently reported.

Current evidence is limited by methodological weaknesses, small sample sizes, and short follow-up periods. Additionally, there is a lack of standardized outcome measures and insufficient representation of older and underserved populations.

Future research should prioritize well-powered, methodologically rigorous trials that examine long-term effects, address equity and accessibility challenges, and explore how digital tools can be sustainably integrated into health systems. Personalized, multicomponent digital approaches may offer the most promise for supporting recovery in this complex and heterogeneous condition.

## Acknowledgments

Generative artificial intelligence tools (ChatGPT, OpenAI) were used in a limited capacity to support language editing and improve clarity in selected sections of the manuscript. All content was reviewed and verified by the authors, who take full responsibility for the final version.

## Funding

The authors declare that no financial support was received for this study.

## Data Availability

All data analyzed in this scoping review are publicly available from the original studies cited in the references. No new datasets were generated or analyzed during this study.

## Authors' Contributions

Conceptualization: SLH, RMB, RGB
Data curation: MVD, VAA
Formal analysis: MVD, VAA
Investigation: SLH, MSC, MPRP, VJF, DH, RA, MVD, VAA, BOB, RMB
Project administration: ALN
Supervision: ALN, CB, RGB
Writing – original draft: SLH, RGB



XSL•FO
RenderX

Writing – review & editing: MSC, MPRP, VJF, DH, RA, MVD, VAA, BOB, RMB

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete search strings for each database.
[DOCX File, 13 KB - i-jmr_v15i1e80616_app1.docx ]

Multimedia Appendix 2
Detailed methodological characteristics and outcome measures for each study.
[DOCX File, 163 KB - i-jmr_v15i1e80616_app2.docx ]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 86 KB - i-jmr_v15i1e80616_app3.docx ]

## References

1. Tsygankova AE, Gerasimov AN, Malov VA, Darvina OV, Volchkova EV, Chulanov VP. Epidemiology, course and predictors of long-term COVID infection (long COVID): a review. Epidemiol Infect Dis 2024;29(1):64-73. [doi: 10.17816/EID610995]
2. Sweis JJ, Alnaimat F, Esparza V, et al. From acute infection to prolonged health consequences: understanding health disparities and economic implications in long COVID worldwide. Int J Environ Res Public Health 2024 Mar 11;21(3):325. [doi: 10.3390/ijerph21030325] [Medline: 38541324]
3. Tabacof L, Tosto-Mancuso J, Wood J, et al. Post-acute COVID-19 syndrome negatively impacts health and wellbeing despite less severe acute infection. medRxiv. Preprint posted online on Nov 6, 2020. [doi: 10.1101/2020.11.04.20226126]
4. Salamanna F, Veronesi F, Martini L, Landini MP, Fini M. Post-COVID-19 syndrome: the persistent symptoms at the post-viral stage of the disease. A systematic review of the current data. Front Med (Lausanne) 2021;8:653516. [doi: 10.3389/fmed.2021.653516] [Medline: 34017846]
5. Vanderlind WM, Rabinovitz BB, Miao IY, et al. A systematic review of neuropsychological and psychiatric sequalae of COVID-19: implications for treatment. Curr Opin Psychiatry 2021 Jul 1;34(4):420-433. [doi: 10.1097/YCO.0000000000000713] [Medline: 34016818]
6. De Luca R, Bonanno M, Calabrò RS. Psychological and cognitive effects of long COVID: a narrative review focusing on the assessment and rehabilitative approach. J Clin Med 2022 Nov 4;11(21):6554. [doi: 10.3390/jcm11216554] [Medline: 36362782]
7. Guo P, Ballesteros AB, Yeung SP, et al. COVCOG 2: cognitive and memory deficits in long COVID: a second publication from the COVID and cognition study. Front Aging Neurosci 2022;14:804937. [doi: 10.3389/fnagi.2022.804937]
8. Panagea E, Messinis L, Petri MC, et al. Neurocognitive impairment in long COVID: a systematic review. Arch Clin Neuropsychol 2025 Jan 21;40(1):125-149. [doi: 10.1093/arclin/acae042] [Medline: 38850628]
9. Rozanski G, Fiorentino A, Hamilton BW, et al. The relationship between cognitive dysfunction and self-reported clinical outcomes in long COVID. Cardiopulm Phys Ther J 2025;36(1):30-38. [doi: 10.1097/CPT.0000000000000271]
10. Skiffington H, Breen C. More than “brain fog”: cognitive dysfunction and the role of occupational therapy in long COVID. Cardiopulm Phys Ther J 2025;36(1):39-49. [doi: 10.1097/CPT.0000000000000274]
11. Kyzar EJ, Purpura LJ, Shah J, Cantos A, Nordvig AS, Yin MT. Anxiety, depression, insomnia, and trauma-related symptoms following COVID-19 infection at long-term follow-up. Brain Behav Immun Health 2021 Oct;16:100315. [doi: 10.1016/j.bbih.2021.100315] [Medline: 34345869]
12. Efstathiou V, Stefanou MI, Demetriou M, et al. Long COVID and neuropsychiatric manifestations (review). Exp Ther Med 2022 May;23(5):363. [doi: 10.3892/etm.2022.11290] [Medline: 35493431]
13. Castanares-Zapatero D, Chalon P, Kohn L, et al. Pathophysiology and mechanism of long COVID: a comprehensive review. Ann Med 2022 Dec;54(1):1473-1487. [doi: 10.1080/07853890.2022.2076901] [Medline: 35594336]
14. Diar Bakerly N, Smith N, Darbyshire JL, et al. Pathophysiological mechanisms in long COVID: a mixed method systematic review. Int J Environ Res Public Health 2024 Apr 12;21(4):473. [doi: 10.3390/ijerph21040473] [Medline: 38673384]
15. Cooper K, Duncan E, Hart-Winks E, et al. Exploring the perceptions and experiences of community rehabilitation for long COVID from the perspectives of Scottish general practitioners’ and people living with long COVID: a qualitative study. BMJ Open 2024 May 15;14(5):e082830. [doi: 10.1136/bmjopen-2023-082830] [Medline: 38749696]
16. Hung P, Brehon K, Miciak M, et al. “I had to know about it, I had to find it, I had to know how to access it”: experiences of access to rehabilitation services among people living with long COVID. Physiother Can 2024;76(1):8-24. [doi: 10.3138/ptc-2023-0021] [Medline: 38465303]



XSL•FO
RenderX

17. Rinn R, Gao L, Schoeneich S, et al. Digital interventions for treating post-COVID or long-COVID symptoms: scoping review. J Med Internet Res 2023 Apr 17;25:e45711. [doi: 10.2196/45711] [Medline: 36943909]
18. Karekla M, Kasinopoulos O, Neto DD, et al. Best practices and recommendations for digital interventions to improve engagement and adherence in chronic illness sufferers. Eur Psychol 2019;24(1):49-67. [doi: 10.1027/1016-9040/a000349]
19. Daniels K, Mourad J, Bonnechère B. Exploring the use of mobile health for the rehabilitation of long COVID patients: a scoping review. Healthcare (Basel) 2024 Feb 10;12(4):451. [doi: 10.3390/healthcare12040451] [Medline: 38391826]
20. Hawke LD, Nguyen AT, Ski CF, Thompson DR, Ma C, Castle D. Interventions for mental health, cognition, and psychological wellbeing in long COVID: a systematic review of registered trials. Psychol Med 2022 Oct;52(13):2426-2440. [doi: 10.1017/S0033291722002203] [Medline: 35768406]
21. Mfouth Kemajou P, Mbanya A, Coppieters Y. Digital approaches in post-COVID healthcare: a systematic review of technological innovations in disease management. Biol Methods Protoc 2024;9(1):bpae070. [doi: 10.1093/biomethods/bpae070] [Medline: 39440031]
22. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
23. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
24. León-Herrera S, Gómez-Bravo R, Sánchez-Castro M, et al. Digital interventions for cognitive and psychological symptoms in long COVID: a scoping review. Open Science Framework. 2024 Dec 11. URL: https://osf.io/hx7ue/overview [accessed 2026-04-22]
25. Amir-Behghadami M, Janati A. Population, Intervention, Comparison, Outcomes and Study (PICOS) design as a framework to formulate eligibility criteria in systematic reviews. Emerg Med J 2020 Jun;37(6):387. [doi: 10.1136/emermed-2020-209567] [Medline: 32253195]
26. Pepa L, Pigliapoco M, Bisoglio P, et al. Therapeutic exercise protocols for people recovering after COVID-19: a tele-health approach. In: Bettelli A, Monteriù A, Gamberini L, editors. Ambient Assisted Living: Springer International Publishing; 2022:355-362. [doi: 10.1007/978-3-031-08838-4_25]
27. Murray E, Goodfellow H, Bindman J, et al. Development, deployment and evaluation of digitally enabled, remote, supported rehabilitation for people with long COVID-19 (living with COVID-19 recovery): protocol for a mixed-methods study. BMJ Open 2022 Feb;12(2):e057408. [doi: 10.1136/bmjopen-2021-057408]
28. León-Herrera S, Oliván-Blázquez B, Sánchez-Recio R, Méndez-López F, Magallón-Botaya R, Sánchez-Arizcuren R. Effectiveness of an online multimodal rehabilitation program in long COVID patients: a randomized clinical trial. Arch Public Health 2024 Sep 18;82(1):159. [doi: 10.1186/s13690-024-01354-w] [Medline: 39294767]
29. Lai CY, Lin CH, Chao TC, et al. Effectiveness of a 12-week telerehabilitation training in people with long COVID: a randomized controlled trial. Ann Phys Rehabil Med 2024 Jun;67(5):101853. [doi: 10.1016/j.rehab.2024.101853] [Medline: 38824899]
30. Hatcher S, Werier J, Edgar NE, et al. Enhancing COVID Rehabilitation with Technology (ECORT): protocol for an open-label, single-site randomized controlled trial evaluating the effectiveness of electronic case management for individuals with persistent COVID-19 symptoms. Trials 2022 Sep 2;23(1):728. [doi: 10.1186/s13063-022-06578-1] [Medline: 36056372]
31. Müllenmeister C, Stoelting A, Schröder D, et al. Evaluating the feasibility, acceptance, and beneficial effects of online occupational therapy for post-COVID-19 condition: protocol for a randomized controlled trial (ErgoLoCo study). JMIR Res Protoc 2024 May 13;13:e50230. [doi: 10.2196/50230] [Medline: 38739435]
32. Smith JL, Deighton K, Innes AQ, et al. Improved clinical outcomes in response to a 12-week blended digital and community-based long-COVID-19 rehabilitation programme. Front Med (Lausanne) 2023;10:1149922. [doi: 10.3389/fmed.2023.1149922] [Medline: 37293307]
33. Al-Jabr H, Windle K, Thompson DR, Jenkins ZM, Castle DJ, Ski CF. Long COVID Optimal Health Program (LC-OHP) to enhance psychological and physical health: protocol for a feasibility randomized controlled trial. JMIR Res Protoc 2022 May 12;11(5):e36673. [doi: 10.2196/36673] [Medline: 35468586]
34. Ennis S, Heine P, Sandhu H, et al. Development of an online intervention for the Rehabilitation Exercise and psycholoGical support After covid-19 InfectioN (REGAIN) trial. NIHR Open Res 2023;3:10. [doi: 10.3310/nihropenres.13371.2] [Medline: 37881468]
35. Vanova M, Patel AM, Scott I, et al. Telehealth-delivered cognitive rehabilitation for people with cognitive impairment as part of the post-COVID syndrome: protocol for a randomised controlled trial as part of the CICERO (Cognitive Impairment in Long Covid: Phenotyping and Rehabilitation) study. Trials 2024 Oct 22;25(1):704. [doi: 10.1186/s13063-024-08554-3] [Medline: 39434179]
36. Widmann CN, Henkel C, Seibert S. "Brain fog" after COVID-19 infection: how the field of neuropsychology can help clear the air. Adv Exp Med Biol 2024;1458:59-76. [doi: 10.1007/978-3-031-61943-4_5] [Medline: 39102190]
37. Cano N, Gómez-Hernández J, Ariza M, et al. A multimodal group-based immersive virtual reality intervention for improving cognition and mental health in patients with post-covid-19 condition. A quasi-experimental design study. Front Psychol 2024;15:1441018. [doi: 10.3389/fpsyg.2024.1441018] [Medline: 39131859]




XSL•FO
RenderX

38. Stavrou VT, Vavougios GD, Kalogiannis P, et al. Breathlessness and exercise with virtual reality system in long-post-coronavirus disease 2019 patients. Front Public Health 2023;11:1115393. [doi: 10.3389/fpubh.2023.1115393] [Medline: 36908417]
39. Groenveld T, Achttien R, Smits M, et al. Feasibility of virtual reality exercises at home for post-COVID-19 condition: cohort study. JMIR Rehabil Assist Technol 2022 Aug 15;9(3):e36836. [doi: 10.2196/36836] [Medline: 35858254]
40. Prudent C, Batt M, Gamraoui S, Danan JL, Klos JY. Télépsychothérapie et COVID-long, une indication envisageable? [Article in French]. Annales Médico-psychologiques, revue psychiatrique 2024 Jun;182(6):520-530. [doi: 10.1016/j.amp.2023.11.012]
41. Krotz A, Sosnowsky-Waschek N, Bechtel S, et al. Reducing sick leave, improving work ability, and quality of life in patients with mild to moderate long COVID through psychosocial, physiotherapeutic, and nutritive supportive digital intervention (MiLoCoDaS): study protocol for a randomized controlled trial. Trials 2023 Dec 8;24(1):798. [doi: 10.1186/s13063-023-07819-7] [Medline: 38066618]
42. Mata R, Bankole AO, Barnhill J, Roth I. A descriptive exploration of younger and older adults' experiences of Integrative Medical Group Visits for long COVID. Aging Health Res 2023 Jun;3(2):100137. [doi: 10.1016/j.ahr.2023.100137] [Medline: 37799164]
43. Montes-Ibarra M, Godziuk K, Thompson RB, et al. Protocol for a pilot study: feasibility of a web-based platform to improve nutrition, mindfulness, and physical function in people living with post COVID-19 condition (BLEND). Methods 2024 Nov;231:186-194. [doi: 10.1016/j.ymeth.2024.10.004] [Medline: 39389403]
44. Tsang W, Tabio L. Feasibility of a manualized long COVID telehealth group therapy within a metropolitan outpatient rehabilitation psychology service. Arch Phys Med Rehabil 2024 Apr;105(4):e141. [doi: 10.1016/j.apmr.2024.02.400]
45. Victoria LW, Oberlin LE, Ilieva IP, et al. A digital intervention for cognitive deficits following COVID-19: a randomized clinical trial. Neuropsychopharmacology 2024 Dec;50(2):472-479. [doi: 10.1038/s41386-024-01995-z] [Medline: 39358543]
46. Eilam-Stock T, George A, Lustberg M, Wolintz R, Krupp LB, Charvet LE. Telehealth transcranial direct current stimulation for recovery from post-acute sequelae of SARS-CoV-2 (PASC). Brain Stimul 2021;14(6):1520-1522. [doi: 10.1016/j.brs.2021.10.381] [Medline: 34655835]
47. Schröder J, Bäuerle A, Jahre LM, et al. Acceptance, drivers, and barriers to use eHealth interventions in patients with post-COVID-19 syndrome for management of post-COVID-19 symptoms: a cross-sectional study. Ther Adv Neurol Disord 2023;16:17562864231175730. [doi: 10.1177/17562864231175730] [Medline: 37255668]
48. Dahmen A, Keller FM, Derksen C, Rinn R, Becker P, Lippke S. Screening and assessment for post-acute COVID-19 syndrome (PACS), guidance by personal pilots and support with individual digital trainings within intersectoral care: a study protocol of a randomized controlled trial. BMC Infect Dis 2022 Aug 15;22(1):693. [doi: 10.1186/s12879-022-07584-z] [Medline: 35971066]
49. Vincent A, Ofovwe O, Gschwandtner M, Shergill S, Faruqui R. Psychiatric, neurophysical and neurocognitive sequelae of post-acute COVID-19 syndrome: a systematic review. BJPsych Open 2022 Jun;8(Suppl 1):S77. [doi: 10.1192/bjo.2022.256]
50. Orfei MD, Porcari DE, D'Arcangelo S, Maggi F, Russignaga D, Ricciardi E. A new look on long-COVID effects: the functional brain fog syndrome. J Clin Med 2022 Sep 21;11(19):5529. [doi: 10.3390/jcm11195529] [Medline: 36233392]
51. Vannorsdall TD, Brigham E, Fawzy A, et al. Cognitive dysfunction, psychiatric distress, and functional decline after COVID-19. J Acad Consult Liaison Psychiatry 2021;63(2):133-143. [doi: 10.1016/j.jaclp.2021.10.006] [Medline: 34793996]
52. Roth A, Chan PS, Jonas W. Addressing the long COVID crisis: integrative health and long COVID. Glob Adv Health Med 2021;10:21649561211056597. [doi: 10.1177/21649561211056597] [Medline: 34820152]
53. Negrut N, Menegas G, Kampioti S, et al. The multisystem impact of long COVID: a comprehensive review. Diagnostics (Basel) 2024 Jan 24;14(3):244. [doi: 10.3390/diagnostics14030244] [Medline: 38337760]
54. Cheng C, Beauchamp A, Elsworth GR, Osborne RH. Applying the electronic health literacy lens: systematic review of electronic health interventions targeted at socially disadvantaged groups. J Med Internet Res 2020 Aug 13;22(8):e18476. [doi: 10.2196/18476] [Medline: 32788144]
55. Mistry SK, Shaw M, Raffan F, et al. Inequity in access and delivery of virtual care interventions: a scoping review. Int J Environ Res Public Health 2022 Aug 1;19(15):9411. [doi: 10.3390/ijerph19159411] [Medline: 35954768]
56. Jensen JD, King AJ, Davis LA, Guntzviller LM. Utilization of internet technology by low-income adults: the role of health literacy, health numeracy, and computer assistance. J Aging Health 2010 Sep;22(6):804-826. [doi: 10.1177/0898264310366161]
57. Miller LM, Callegari RA, Abah T, Fann H. Digital literacy training for low-income older adults through undergraduate community-engaged learning: single-group pretest-posttest study. JMIR Aging 2024 May 14;7:e51675. [doi: 10.2196/51675] [Medline: 38599620]
58. Borghouts J, Eikey E, Mark G, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387. [doi: 10.2196/24387] [Medline: 33759801]
59. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
60. Timpel P, Harst L. Research implications for future telemedicine studies and innovations. Eur J Public Health 2019 Nov;29(Supplement_4). [doi: 10.1093/eurpub/ckz185.098]




XSL•FO
RenderX

61. Pirosca S, Clarke M, Treweek S. Learning from Cochrane systematic reviews: what improvements do these suggest for the design of trials? F1000Res 2020 Mar 11;9:178. [doi: 10.12688/f1000research.22635.2]

## Abbreviations

**CBT:** cognitive behavioral therapy
**mHealth:** mobile health
**PRISMA-ScR:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
**VR:** virtual reality

*Edited by M Balcarras; submitted 14.Jul.2025; peer-reviewed by DL Gustafson, T Keller; revised version received 14.Apr.2026; accepted 14.Apr.2026; published 08.Jun.2026.*

*Please cite as:*
*León-Herrera S, Sánchez-Castro M, Luisa Neves A, Rodríguez-Pérez MP, Jobim Fischer V, Hutmacher D, Aldakhil R, Vaillancourt de Dios M, Anjos de Almeida V, Oliván-Blázquez B, Magallón-Botaya R, Benoy C, Gómez-Bravo R*
*Digital Interventions Addressing Cognitive and Psychological Symptoms in Long COVID: Scoping Review of Multicomponent Approaches*
*Interact J Med Res 2026;15:e80616*
*URL: https://www.i-jmr.org/2026/1/e80616*
*doi:10.2196/80616*

© Sandra León-Herrera, Marta Sánchez-Castro, Ana Luisa-Neves, Mª Pilar Rodríguez-Pérez, Vinicius Jobim Fischer, Djenna Hutmacher, Reham Aldakhil, Marina Vaillancourt de Dios, Vinicius Anjos de Almeida, Bárbara Oliván-Blázquez, Rosa Magallón-Botaya, Charles Benoy, Raquel Gómez-Bravo. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 8.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Pulse Discovery Toolkit, a Multicomponent Nutrition Intervention for Preschool Children in Childcare Centers: Mixed Methods Pilot Study

Hiwot Abebe Haileslassie[1,2], PhD; Renee Ramikie[1], MSc; Hassan Vatanparast[1,3], MD, PhD; D Dan Ramdath[1], PhD; Amanda Froehlich Chow[3], PhD; Phyllis Shand[4], PhD; Rachel Engler-Stringer[5], PhD; Jessica R L Lieffers[1], PhD; Shannon Hood-Niefer[6], PhD; Carol Henry[1], PhD

[1]College of Pharmacy and Nutrition, University of Saskatchewan, Saskatoon, SK, Canada

[2]Agricultural, Food and Nutritional Science, University of Alberta, 8602-112 St, Edmonton, AB, Canada

[3]School of Public Health, University of Saskatchewan, Saskatoon, Saskatoon, SK, Canada

[4]College of Agriculture and Bioresources, University of Saskatchewan, Saskatoon, SK, Canada

[5]Department of Community Health and Epidemiology, University of Saskatchewan, Saskatoon, SK, Canada

[6]Saskatchewan Food Industry Development Centre Inc, Saskatoon, SK, Canada

**Corresponding Author:**
Hiwot Abebe Haileslassie, PhD
College of Pharmacy and Nutrition, University of Saskatchewan, Saskatoon, SK, Canada

## Abstract

**Background:** Children's eating habits are formed at an early age, making childhood a crucial period for introducing novel foods, such as pulse-based food products. Pulse Discovery Toolkit (PDTK) intervention was designed to increase familiarity with pulses and to eventually contribute to the consumption of pulse-based foods among preschool children in childcare centers (CCs).

**Objective:** To determine PDTK's impact on knowledge, acceptability, and consumption of pulse-based foods among preschool children attending CCs, and to assess its feasibility and acceptability by early childhood educators (ECE) and cooks. The nutrient contents and food group servings of pulse-based intervention recipes in the PDTK were also compared with regular CC recipes.

**Method:** The PDTK intervention was delivered over a 3-month period in 2 CCs in Saskatoon (50 children, 8 staff). The intervention, which integrated taste exposure and nutrition education, consisted of 12 child-friendly weekly lessons, a food service guide for cooks, 15 recipes for pulse-based foods, 4 intervention recipes incorporated in the CC menu, and 4 parent newsletters. Mixed methods were used with pre- and postintervention knowledge tests, plate waste measurement, sensory evaluation, ECE and cook's perspective, and nutrient content comparison of the intervention and control foods from the regular childcare menu to evaluate the intervention's impact.

**Result:** Improvements in correct identification of chickpeas (2/21 [10%] at preintervention to 7/21 [33%] at postintervention, *P*=.074), beans (8/21 [38%] to 11/21 [52%], *P*=.68), and peas (6/21 [27%] to 8/21 [38%], *P*=.61) were not statistically significant. Children consumed higher amounts of the regular recipes (293.54, SD 27.65; 178.46, SD 24.33) than the intervention recipes (211.56, SD 25.61; 108.83, SD 21.97) at both times, respectively. However, at the end of the intervention, significant differences were only observed in the amount of total food consumption (*P*=.049) and the protein content (*P*=.04) when consumption proportion was examined, with both being higher for the control recipes in comparison to the intervention recipes. The majority (92% and 72%) of the children rated the refried bean wrap and lentil smoothie, "yummy," respectively. Most of the intervention recipes have lower energy, fat, and sodium content compared with the regular CC recipes. Findings from ECE semistructured interviews and the lesson plan evaluations revealed that the ECEs reacted favorably to the curriculum. The cooks from the participating CCs did not report any barriers to cooking pulses in their facility. However, the need for modification to make the recipes easier to cook in CCs was noted in our study.

**Conclusions:** With a few modifications to make some of the lessons more age-appropriate and some of the recipes easier to cook, it is feasible to implement the PDTK in CCs in order to promote regular consumption of pulses.

**International Registered Report Identifier (IRRID):** RR2-10.2196/22775

***(Interact J Med Res 2026;15:e55866)*** doi:10.2196/55866

**KEYWORDS**

childcare center; nutrition intervention; preschool children; pulse; sensory evaluation; pediatric; child; children; nutrition; nutritional; dietary



XSL•FO
RenderX

## Introduction

Children's eating habits are formed at an early age, making the first 5 years of life a crucial period for introducing novel healthy foods. A growing body of evidence supports the need for repeated exposure to ensure liking and preference of novel foods [1-3]. Many researchers contend that childcare centers (CCs) are useful for the delivery of nutrition interventions as they provide continuous and intensive contact with children at an age when they develop life-long dietary preferences and eating patterns [4]. The fact that children often consume about 70% of their daily nutrient intake in this environment [5] and 52% of children younger than 6 years old attend CCs increases the significance of childcare institutions for nutrition intervention.

Sustainability, cost, nutrition, and health value make pulse-based foods (foods made with beans, chickpeas, peas, and lentils) a suitable dietary choice for establishing healthy eating habits among preschool children in CCs. Young children who consumed beans were found to have greater protein, fiber, potassium, and magnesium intake compared with those who did not [6]. Despite these benefits, the consumption of pulse-based foods is low in Canada [7]. Further, Canada is a global leader in pulse crop production. An increase in awareness through repeated exposure and skill in cooking pulses can potentially lead to liking of pulse-based dishes and more pulse-based dishes being offered in CCs [1,8], which can, in turn, lead to an increase in the consumption of these foods.

While much effort has been dedicated to improving fruit and vegetable intake in CCs [9-12], interventions promoting pulse-based dishes in CCs are scarce. Built on social cognitive theory and taking Piaget's cognitive development into consideration, the Pulse Discovery Toolkit (PDTK) intervention was developed to promote consumption of pulse-based foods in CCs [13]. The PDTK is designed to offer multiple opportunities for children to taste and learn about pulses either through lessons or through pulse-based snacks and lunches. This pilot study aimed to compare knowledge, willingness to taste, and consume pulse-based foods in preschool children in 2 CCs pre- and post-PDTK intervention. In addition, nutrient contents and food group servings of the pulse-based intervention recipes were compared with the regular CC recipes. The feasibility and acceptability of the intervention from early childhood educators' (ECE's) and cooks' perspectives were also assessed.

## Method

### Overview

A pre-post design pilot study with a community-based approach was carried out in 2 CCs in Saskatoon. ECEs, staff, and children in CCs were engaged in revising the PDTK before its rollout. A detailed description of the method is published by Haileslassie et al [13]; this study used a multimethod approach with both quantitative and qualitative data collection.

### Ethical Consideration

Ethical approval was obtained from the University of Saskatchewan Behavioral Research Ethics Board, and a certificate of approval was issued (Beh # 15 - 388). Prior to data collection, consent and verbal assent were obtained from parents and children, respectively. As a pilot study designed to assess feasibility, 50 children were recruited from 2 CCs, a sample size consistent with recommendations in Teresi et al [14].

### The Intervention and Study Population

The PDTK was developed using principles of intervention mapping, and it was delivered over the course of 3 months in 2 CCs in Saskatoon (50 children, 8 staff) in 2016. Both centers are located in the city and provide food prepared on-site by their cook. The children in these centers are offered a morning snack, lunch, and an afternoon snack. Prior to implementation, each of the 2 facilities received copies of the PDTK manual, and ECEs and cooks received training on using the manual and the overall objectives of the project. The PDTK manual consisted of 12 lesson plans, a food service guide, 15 recipes, and 4 parent newsletters. One of the participating CCs was English-speaking, and one was French-language only. The latter required the PDTK manual to be translated into French. Trained graduate nutrition students and ECEs participated in the delivery of the lessons.

### Knowledge Assessment Questionnaire

Pre- and posttest assessment of the children's pulse knowledge was carried out at each intervention site. A pictorial data collection questionnaire was adapted and pretested for this purpose. A similar type of pictorial tool has been previously validated and used with children [15,16].

### Plate Waste Assessment and Comparison of Nutrient Composition and Food Group Servings of Intervention and Control Recipes

The 12 weeks of the intervention were divided into 3 cycles: time-point 1 (the first 4-wk cycle), adjustment period, and time-point 2 (the last 4-wk cycle). Recipes that could possibly serve as a control for the intervention recipes were selected from the regular recipes offered in each CC (Table 1). The matching of the intervention recipes was based on their similarity to control menu items, likelihood of adaptability, and cooking time. In addition to the pulse-based foods prepared at CCs, commercially purchased products, such as puffs (extruded light and airy snacks), lentil pasta, and biscuits, were also evaluated by the children. The puff flavors were particularly developed in consultation with food science experts, educators, and children. These foods were incorporated to increase plant-based snack and meal options as opposed to the regular high-salt processed foods.



XSL•FO
RenderX

**Table .** Intervention recipes and their corresponding controls in the participating childcare centers in Saskatoon.

| Intervention recipes | Control recipes |
|---|---|
| Lentil pizza | • Mexican Pizza<br>• Chicken Pizza |
| Chickpea spread | • Hummus<br>• Cream Cheese |
| Stir fry chickpea, vegetables | • Stir Fry Vegetables with Beef<br>• Kidney Beans, Beef, and Vegetables |
| Three bean quesadillas | • Chicken and Cheese Quesadillas<br>• Tuna Wrap |

Plate waste measurement of the intervention and the regular recipes was taken twice during the 12-week intervention. A digital photography-enhanced application validated in previous studies captured the weight of the foods before and after consumption [17]. Food Processor Nutritional Analysis Software (Esha version 10.10.00) was used to compute the energy, nutrient breakdown (protein, total fiber, fat, saturated fat, calcium, potassium, sodium, and iron), and food group servings (meat and alternative, milk, grain, vegetable, and fruit) in order to compare the nutrient composition and food group servings of intervention and control recipes.

### Sensory Evaluation

Three types of sensory evaluation were carried out to assess the acceptability of the pulse-based intervention recipes. First, sensory evaluation was carried out as part of the lesson on 3 occasions. The children were asked to taste and report their liking using a 3-point facial hedonic scale with the options of "Yummy," "Yucky," and "Ok" [18]. Second, sensory evaluation of selected recipes was conducted at 2 time points to compare the degree of liking after exposure. The first was closer to the beginning of the intervention, and the second was taken at the end after several exposures to pulse-based dishes. Third, preference tasting was carried out by presenting 2 pulse-based foods at the same time and requesting the children to indicate which of the two they like the most. This was particularly used for tasting spreads, and participants were offered a choice of trying it with and without accompanying vegetables. Further details are available in the protocol by Haileslassie et al [13].

### Qualitative Data Collection: Lesson Plan Evaluation, ECE, and Cook Interviews

Weekly lesson plan evaluation using questions adapted from Sharma et al [9] was carried out at the end of each lesson. One-on-one semistructured interview guides were used to gather information from cooks and ECEs. A semistructured postintervention interview adapted from Sharma et al [19] was used to further explore the acceptability and feasibility of the PDTK educational resource by the ECEs. The cooks' interview focused on the perceived benefits and barriers of cooking pulses and the feasibility of cooking the recipes included in the PDTK. Interviews were recorded, transcribed, and analyzed into themes [20]. The participants were given a chance to review their transcripts to validate their views or opinions.

### Data Analysis and Quality Control

To ensure a standardized procedure was followed for data collection, research assistants were trained on administering the pre-post questionnaire, conducting sensory evaluation, and plate waste measurement. Furthermore, each assistant was trained on how to instruct and prepare the participating children for sensory evaluation. A one-day training session, along with practice in a pretest CCs using data not included in the final analysis, was conducted to ensure that research assistants fully understood the data collection protocols and were proficient in applying them.

Quantitative data were analyzed using SPSS (version 24; IBM Corp; Armonk, NY, USA) and SAS (version 9.4; SAS Institute Inc. Cary, NC, USA). Results were considered statistically significant if a $P$ value <.05 was obtained. SPSS was used to generate descriptive statistics and for paired comparisons. Each item on the Knowledge Assessment Questionnaire was scored as a binary categorical variable (1=correct identification; 0=incorrect identification). For each pulse item, we calculated the proportion of children answering correctly at pre- and postintervention. Chi-square and McNemar tests were used for paired nominal data to compare the pre and post pulse knowledge scores. A paired sample $t$ test was also conducted on spreads (green split pea vs red bean) to determine if there was any difference between sensory acceptance at the beginning and at the end of the intervention. SAS was used to generate $t$ test and generalized estimating equation (GEE) model results for plate waste measurements. The GEE model was developed to analyze the relationship between control recipes and intervention recipes, along with the co-covariates age, sex, and center across the 2 time periods (baseline and endpoint) using aggregate data. The outcome variable was the amount of food consumed, and a fixed effect was selected. The independent variables included the type of recipe (intervention vs control), time (baseline and endpoint), and covariates (age, sex, and center). The GEE model analysis is a widely used statistical model for data collection for repeated measurements, as it allows analysis for missing and unbalanced data. Since the distribution of all responses (outcomes) is from the gamma distribution family, we were required to choose gamma distribution and "power (-1)" as link function for conducting "pro Genmod" in SAS. While the GEE model accounts for missing data, a sensitivity analysis was not conducted. Using aggregate data, an independent $t$ test was performed to compare the mean



XSL•FO
RenderX

consumption and the mean consumption proportion of both the intervention and control recipes. Consumption proportion (proportion of food consumed to food served) was calculated for each intervention and control recipe using the formula provided below:

Consumption proportion = (weight of food served − weight of food leftover) / weight of food served

The qualitative data analysis followed thematic analysis, consistent with the coding reliability approach described by Braun and Clarke [21]. The validated transcripts from ECE and cook interviews were analyzed separately to reflect their distinct perspectives. In the first phase, transcripts were reviewed line by line to identify recurring ideas and patterns. In the second phase, initial codes were independently reviewed by a second researcher to verify consistency and refine theme development. In the third phase, emerging themes were discussed collaboratively to ensure coherence and resolve discrepancies through consensus. Relevant participant quotes were organized under final thematic categories to illustrate and support the findings. To enhance trustworthiness and rigor, the team used researcher triangulation, peer debriefing, and an audit process. Themes were refined iteratively to ensure they accurately represented the data.

### Ethical Consideration

Ethical approval was obtained from the University of Saskatchewan Behavioral Research Ethics Board. Prior to data collection, consent and verbal assent were obtained from parents and children, respectively.

## Result

### Overview

A total of 50 children between the ages of 2 and 5 years from 2 CCs participated in the study. All ECEs (n=5) from both schools completed the lesson plan evaluation forms. In addition, the ECEs and cooks in each center participated in a postintervention interview.

### Pre and PostIntervention Pulse Knowledge of Preschool Children

Pre and postintervention pulse knowledge of the preschool children is presented in Table 2. No statistically significant improvements in children's pulse knowledge from pre to postintervention across all items were noted, though the increase in correct identification of chickpeas approached significance ($P$=.07).

**Table .** Pre and postintervention pulse knowledge of preschool children in 2 childcare centers in Saskatoon (n=21). The number of children who have a complete set of pre-post knowledge assessment data in a McNemar's test.

| Questions | Pretest, n (%) | | Posttest, n (%) | | *P* value |
|---|---|---|---|---|---|
| | Correct | Incorrect | Correct | Incorrect | |
| Distinguishing between lentils and beans | 14 (67) | 7 (33) | 13 (62) | 8 (38) | ≥.99 |
| Identification of chick-peas | 2 (10) | 19 (91) | 7 (33) | 14 (67) | .07 |
| Identification of lentils | 1 (5) | 20 (95) | 2 (10) | 19 (90) | ≥.99 |
| Identification of beans | 8 (38) | 13 (62) | 11 (52) | 10 (48) | .68 |
| Identification of peas | 6 (27) | 15 (71) | 8 (38) | 13 (62) | .61 |

### Plate Waste Measurement of Pulse-Based Foods

The amount of food consumed from the intervention and the control recipes at time-point 1 and time-point 2 is described in Table 3. Significant differences in mean consumption between the control and intervention recipes were noted among most of the variables at both times, except for the fruit and vegetable servings, meat and alternative servings, and fiber content. The GEE Model analysis showed significantly lower mean consumptions in time-point 2 for all the variables of interest, except for the fruit and vegetable servings. The mean consumption proportion from the intervention and the control recipes at time-point 1 and time-point 2 is presented in Table 4. At time-point 1, no significant difference was observed in the mean consumption proportions of all the nutrients of interest between the intervention and control recipes, except for the higher saturated fat from the control recipes and the higher trans-fat from the intervention recipes. At time-point 2, significant differences were only observed between the mean consumption proportion of the quantity ($P$=.049) and the protein content ($P$=.04); both of which were higher among the control recipes in comparison to the intervention recipes. In terms of the consumption proportion, the GEE model revealed significant differences in the total quantity and most nutrients, except for the servings of fruit and vegetables, meat, whole grain, trans-fat, and sugar content.



XSL•FO
RenderX

**Table .** The quantity, food groups serving, and nutrient contents of intervention and control recipes consumed by preschool children in 2 childcare centers at time-point 1 and time-point 2 during a 12-week pulse discovery tool kit intervention.

| | Time-point 1 | | | | Time-point 2 | | | | G EEA |
|---|---|---|---|---|---|---|---|---|---|
| | Control (n=39), mean (SD) | Intervention (n=38), mean (SD) | Mean difference[b] | *P* value[c] | Control (n=37), mean (SD) | Intervention (n=27), mean (SD) | Mean Difference[b] | *P* value[c] | *P* value |
| Quantity (g) | 293.5 (172.7) | 211.6 (157.9) | 81.97 | .03 | 178.5 (148.0) | 108.8 (114.2) | 69.63 | .049 | <.001 |
| Fruit and vegetables serving | 0.95 (0.95) | 0.96 (0.76) | −0.01 | .95 | 0.42 (0.49) | 0.46 (0.61) | −0.04 | .77 | .21 |
| Meat and alternative serving | 0.98 (0.54) | 0.74 (0.59) | 0.24 | .06 | 0.62 (0.50) | 0.40 (0.44) | 0.22 | .06 | .001 |
| Whole grain serving | 39.10 (25.11) | 21.55 (20.78) | 17.54 | <.001 | 28.96 (24.92) | 4.67 (8.40) | 24.28 | <.001 | <.001 |
| Protein (g) | 30.82 (15.61) | 20.24 (14.79) | 10.58 | .003 | 21.64 (15.61) | 9.50 (10.25) | 12.13 | <.001 | <.001 |
| Fiber (g) | 5.92 (3.40) | 5.24 (4.00) | 0.68 | .42 | 3.79 (3.49) | 2.16 (2.32) | 1.62 | .04 | .02 |
| Carbohydrates (g) | 56.50 (38.75) | 36.02 (27.27) | 20.48 | .01 | 38.94 (37.90) | 19.03 (17.89) | 19.91 | .02 | <.001 |
| Calories (kcals) | 554.3 (303.5) | 316.8 (236.6) | 237.5 | <.001 | 387.4 (296.6) | 155.1 (152.3) | 232.2 | <.001 | <.001 |
| Sodium (mg) | 1255.1 (882.9) | 531.0 (405.8) | 724.1 | <.001 | 624.5 (499.5) | 285.5 (276.7) | 339.0 | .002 | <.001 |
| Saturated fat (g) | 9.18 (4.75) | 4.73 (4.06) | 4.45 | <.001 | 6.60 (4.51) | 2.05 (2.64) | 4.55 | <.001 | <.001 |
| Trans fat (g) | 0.32 (0.35) | 0.005 (0.01) | 0.31 | <.001 | 0.26 (0.30) | 0 | 0.26 | <.001 | .01 |
| Sugar (g) | 4.38 (2.79) | 3.18 (2.39) | 1.20 | .049 | 2.85 (2.78) | 1.56 (1.67) | 1.28 | .04 | <.001 |

[a]GEE: generalized estimating equation.

[b]Note-mean differences: subtracting means of control variables from means of intervention variables.

[c]Independent *t* test.



XSL•FO
RenderX

**Table .** The consumption proportion (Consumption proportion=sum of the amount of all intervention (or control foods) a specific child eats sum of the amount of intervention foods (or control foods) that was given to that specific child) of intervention and control recipes at time-point 1 and time-point 2 during a 12-week pulse discovery toolkit intervention in 2 childcare centers in Saskatoon.

| | Time-point 1 | | | | Time-point 2 | | | | GEE[a] |
|---|---|---|---|---|---|---|---|---|---|
| | Control (n=39), mean (SD) | Intervention (n=38), mean (SD) | Mean difference[b] | *P* value[c] | Control (n=37), mean (SD) | Intervention (n=27), mean (SD) | Mean difference[b] | *P* value[c] | *P* value |
| Quantity (g) | 0.79 (0.23) | 0.73 (0.28) | 0.06 | .27 | 0.86 (0.20) | 0.73 (0.30) | 0.13 | .05 | .001 |
| Fruit and vegetables servings | 0.78 (0.25) | 0.74 (0.29)[d] | 0.04 | .56 | 0.81 (0.21)[e] | 0.68 (0.35) | 0.13 | .09 | .08 |
| Meat servings | 0.76 (0.27) | 0.74 (0.27) | 0.02 | .77 | 0.82 (0.24) | 0.73 (0.30) | 0.10 | .14 | .09 |
| Whole grain servings | 0.74 (0.33) | 0.68 (0.37)[f] | 0.06 | .51 | 0.82 (0.29) | 0.87 (0.29) | −0.05 | .65 | .08 |
| Protein (g) | 0.80 (0.23) | 0.73 (0.28) | 0.07 | .28 | 0.85 (0.21) | 0.72 (0.30) | 0.13 | .04 | <.001 |
| Fiber (g) | 0.78 (0.26) | 0.70 (0.31) | 0.08 | .24 | 0.85 (0.21) | 0.73 (0.31) | 0.12 | .07 | .003 |
| Carbohydrates (g) | 0.81 (0.22) | 0.71 (0.29) | 0.09 | .13 | 0.87 (0.21) | 0.76 (0.30) | 0.11 | .09 | .002 |
| Calories (kcals) | 0.80 (0.22) | 0.71 (0.29) | 0.09 | .15 | 0.86 (0.20) | 0.74 (0.29) | 0.12 | .07 | <.001 |
| Sodium (mg) | 0.79 (0.22) | 0.72 (0.29) | 0.07 | .23 | 0.86 (0.20) | 0.75 (0.29) | 0.11 | .09 | .005 |
| Saturated Fat (g) | 0.82 (0.19) | 0.67 (0.3) | 0.15 | .02 | 0.86 (0.20) | 0.75 (0.31) | 0.12 | .08 | <.001 |
| Trans fat (g) | 0.76 (0.30) | 1.00 (0)[g] | −0.24 | .03 | 0.85 (0.26) | __[h] | — | — | .22 |
| Sugar (g) | 0.78 (0.24) | 0.72 (0.29) | 0.07 | .25 | 0.85 (0.21) | 0.73 (0.31) | 0.12 | .06 | .001 |

[a]GEE: generalized estimating equation.

[b]Mean differences: subtracting means of control variables from means of intervention variables.

[c]Independent *t* test.

[d]Intervention recipes (n=36).

[e]Control (n=28) and intervention recipes (n=26).

[f]Intervention recipes (n=34).

[g]Intervention recipes (n=8).

[h]Not applicable.

## Sensory Evaluation of Pulse-Based Foods

The results obtained from sensory evaluation of lentil pasta, refried bean wrap, lentil smoothie, and 2 types of extruded snacks (puffs) are presented in Table 5. Majority (92% and 72%) of the children rated the refried bean wrap and lentil smoothie “yummy,” respectively. Some children (9.4%, 17%, and 15%) refused to taste the lentil smoothie, red pepper and lime flavor puff, and white cheddar puff, respectively.



XSL•FO
RenderX

**Table .** Sensory evaluation of selected pulse-based foods by preschool children during pulse discovery toolkit intervention in 2 childcare centers in Saskatoon.

| Pulse-Based foods | Hedonic rating[a] | | | |
|---|---|---|---|---|
| | Yummy, n (%) | OK, n (%) | Yucky, n (%) | Not willing to taste, n (%) |
| Lentil pasta (n=17) | 10 (59) | 4 (24) | 3 (17.6) | —[b] |
| Refried bean wrap (n=25) | 23 (92) | 1 (4) | 1 (4) | — |
| Lentil smoothie (n=32) | 23 (72) | 2 (6) | 4 (12.5) | 3 (9.4) |
| Red pepper and lime flavor puff (n=12) | 8 (67) | 0 (0) | 2 (17) | 2 (17) |
| White cheddar flavor puff (n=13) | 6 (46) | 4 (31) | 1 (8) | 2 (15) |

[a]A 3-point facial hedonic scale with ratings of “Yummy,” “Yucky,” and “Ok”.

[b]Not applicable.

## Comparison of the Nutrient Composition Between Intervention and Control Recipes

The nutrient composition of intervention and control recipes is provided in Table 6. The comparison revealed higher energy, lower fat, and lower sodium content in most of the intervention recipes.



XSL•FO
RenderX

**Table .** The nutrient composition of 100 g of the pulse discovery toolkit intervention and the control recipes from 2 childcare centers in Saskatoon.

| Recipes | Energy (Kcals) | Protein (g) | Total fiber (g) | Fat (g) | Saturated fat (g) | Ca (mg) | K (mg) | Na (mg) | Iron (mg) | Meat svgs[a] | Milk svgs | Grain svgs | Veg/fr[b] svgs |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mexican pizza (R)[c] | 134.58 | 8.17 | 4.22 | 3.04 | 1.3 | 190.02 | 285.44 | 460.52 | 1.59 | 0.1 | 0.22 | 0.63 | 0.16 |
| Lentil pizza (I) | 141.64 | 8.84 | 2.86 | 5.87 | 3.34 | 143.1 | 254.47 | 186.46 | 1.19 | 0.21 | 0.21 | 0.20 | 0.40 |
| Chicken pizza (R)[d] | 206.69 | 10.37 | 1.6 | 6.19 | 3.25 | 148.88 | 197.78 | 362.92 | 1.61 | 0.10 | 0.30 | 1.22 | 0.39 |
| Lentil pizza (I) | 142.83 | 8.78 | 2.55 | 5.64 | 3.29 | 144.17 | 247.19 | 211.21 | 1.71 | 0.21 | 0.40 | 0.21 | 0.40 |
| Hummus and tortilla (R) | 245.92 | 8.68 | 4.36 | 3.04 | 1.3 | 78.92 | 34.58 | 474.99 | 2.27 | 0.30 | 0 | 1.90 | 0 |
| Chickpea spread and tortilla (I) | 243.3 | 8.61 | 4.42 | 5.24 | 3.34 | 80.16 | 39.38 | 503.57 | 2.28 | 0.30 | 0 | 1.90 | 0.03 |
| Pita with cream cheese (R) | 297.03 | 8.06 | 1.48 | 12.07 | 6.46 | 89.95 | 125.92 | 465.3 | 1.88 | 0.30 | 0 | 1.90 | 0 |
| Pita with chickpea spread (I) | 171.13 | 6.84 | 0.67 | 3.23 | 0.37 | 53.64 | 117.17 | 453.52 | 1.33 | 0.30 | 0 | 1.90 | 0.03 |
| Stir fry vegetables with beef (R) | 202.12 | 14.58 | 0.42 | 14.51 | 4.45 | 22.51 | 257.86 | 309.5 | 1.15 | 0.30 | 0 | 1.90 | 0 |
| Stir fry chickpea, vegetables, and beef (I) | 109.77 | 11.45 | 1.27 | 4.43 | 1.42 | 27.56 | 253.61 | 145.53 | 1.35 | 0.30 | 0 | 1.90 | 0.03 |
| Kidney beans, beef, and vegetable (R) | 136.13 | 12.09 | 1.94 | 6.62 | 2.34 | 32.97 | 278.07 | 145.15 | 1.48 | 0.72 | 0 | 0 | 0.34 |
| Stir fried chicken, chickpea, and vegetable (I) | 86.32 | 10.7 | 1.27 | 1.87 | 0.27 | 29.36 | 238.72 | 131.19 | 0.78 | 0.50 | 0 | 0 | 0.96 |



XSL•FO
RenderX

| Recipes | Energy (Kcals) | Protein (g) | Total fiber (g) | Fat (g) | Saturated fat (g) | Ca (mg) | K (mg) | Na (mg) | Iron (mg) | Meat svgs[a] | Milk svgs | Grain svgs | Veg/fr[b] svgs |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chicken and cheese quesadillas (R) | 273.02 | 20.64 | 1.55 | 15.74 | 8.33 | 270.87 | 132.39 | 375.94 | 1.4 | 0.53 | 0.70 | 0.70 | 0 |
| Three bean quesadillas | 231.69 | 10.63 | 4.46 | 9.37 | 4.04 | 166.72 | 132.06 | 353.46 | 2.06 | 0.30 | 0.31 | 1.09 | 0.10 |
| Tuna wrap (R) | 228.53 | 10.87 | 3.72 | 8.15 | 1.65 | 53.56 | 78.03 | 422.7 | 1.87 | 0.30 | 0 | 1.45 | 0.21 |
| Three bean quesadillas (I) | 219.54 | 10.89 | 4.64 | 8.07 | 3.92 | 162.67 | 132.06 | 364.57 | 2.14 | 0.30 | 0.31 | 1.09 | 0.10 |

[a]Svgs: servings.

[b]Veg/fr: vegetable and fruit.

[c]*R*: Regular Recipe.

[d]I: Intervention recipe.

## Qualitative Findings: Educator and Cook Perspectives on the PDTK Intervention

The qualitative data are presented in three distinct sections: (1) open-ended lesson plan evaluations, (2) ECE interviews, and (3) cook interviews. These are presented separately to reflect the distinct data collection and analysis processes used for each group.

### Lesson Plan Evaluations

ECEs rated the clarity of the instructions of most lessons as "good" on a 5-point scale that ranges from poor, fair, good, very good, to excellent. They indicated that most of the learning objectives were appropriate for this age group. However, ECEs also reported that some of the activities were difficult for the children in this age group to comprehend. One of the ECEs mentioned particularly in lesson 12 (mystery bucket).

> *I think it was difficult for children to identify food groups.*

The ECEs also noted children's reactions to the lessons and reported that most children were engaged in the PDTK activities. Examples of responses are as follows:

> *They had smiles on their faces as they were eager to answer the questions correctly to show what they already knew.* [Lesson 1- Healthy Eating]
>
> *They were engaged and enthusiastic and seemed to have fun.* [Lesson 2- Meet the pulses]
>
> *Dramatic play is one of the children's favorite games. They couldn't wait to have a turn being a server, customer or cook.* [Lesson 10- Let's Play Restaurant]
>
> *They loved it. They concentrated on their menus and had a lot of fun playing restaurant.* [Lesson 10 - Let's Play Restaurant]

Recommendations on improving the acceptability and feasibility of these lesson plans included breaking down the lesson plan activities into smaller components, as well as dividing the children into smaller groups to reduce wait times and the risk of loss of interest. The following quotes are responses related to the acceptability of PDTK activities for implementation in their CCs.

> *Yes, anything promoting active lifestyle is always acceptable.* [Lesson 9 Pulse Bowling]
>
> *Yes, having children be part of the meal preparation encourages healthy eating.* [Lesson 10 - Let's Make Our Own Food]
>
> *Yes, we can insert [the program] into the curriculum for the health of the Canadian children and to avoid obesity in children.* [Lesson 11- Let's play restaurant]

### Educators' Perception of the PDTK

Key themes were collated from ECE responses about their PDTK experiences. The primary themes identified from the interviews were as follows: (1) feasibility and acceptability of the PDTK, (2) barriers to implementation, and (3) recommendations for the PDTK. Table 7 provides a snapshot of themes developed and examples of corresponding quotes from the educators' observations and experiences throughout the intervention.



XSL•FO
RenderX

**Table .** A snapshot of the themes and corresponding quotes identified from interviews.

| Themes | Quotes |
|---|---|
| 1.Feasibility and acceptability of the PDTK | • *Since you guys started, I found every day was a positive influence on the kids. I really enjoyed that you guys came and did these. We don't often have the opportunity to have something like this to encourage healthy eating. We do it on our own and make up our own. So, to have someone come in that already has something set up and in place was awesome, because I love encouraging kids to eat healthy food.* [ECE # 1]<br>• *For this place, the people are not moving too much and the kids are not moving too much. It is not like, I came from [name of country] and the kids they can walk and we don't have a problem with winter. They stay at home and they are not moving. Sometimes I see what there are eating too much is giving them obesity. But, when you have food that control it. Like your program it will help them to stay healthy. It will help them. I would like to put it that way. If I had my own center, like this I would put in my own program as quickly.* [ECE #2]<br>• *I think we can try some activities. So, for example, we made taco with beans and I think it is a good curriculum for children. Because, the activities included sensory activities. So we could taste it, smell it and feel. All sensory were included in these activities. Also, children they are interested in making food and touching something. So, we can do them next time.* [ECE #3]<br>• *They were engaged because sometimes they would ask normally when you people are coming…even the kids would ask, what they are going to give us to eat today. They will ask like that or maybe they will ask … (the cook)…. what are you going to cook for us today? Beans and what, they need to know.* [ECE #3] |
| 2.Barriers to implementation | • *Not, that I didn't enjoy that is not the word. I think some activities were very, not difficult but were not really appropriate for their age, is appropriate the word? (Yea). They were very young for the kind of activities. It wasn't really for them. It wasn't really helpful.* [ECE #1]<br>• *don't think it's important to know about food groups and guide. We don't need for children. We just can explain that milk is for teeth and meat is for our health and just explain simply and they can accept. If we added some information it it's for first group and some food include in the first group….. They are just 2 and half years old, it's difficult for them.* [ECE #2]<br>• *Actually, we didn't have time to do the activity. We have 20 children and plus we have two teachers with the 20 children. For me I think it was difficult to do these lessons. Because, before starting these activities we need to explain what we are going to do and what kind of things I do with some stuff. But, actually, sometimes when I was alone with many children it was so difficult.* [ECE #3] |
| 3. Recommendation for the PDTK | • *What I enjoyed was when they had to see the different kind of peas, the different colors and when we had to do the smoothies. Because they have to count, measure and they have to taste, they have to touch. It was really physical; you know sensory for them and when they have to grow all the seeds.* [ECE #4]<br>• *We need more big charts, big pictures, not just the small paper to show exactly what it actually is. Also, we need, more time to explain what it is. So, I prefer, one day we can do about the first food group and the next day the second group ….…..…* [ECE #5] |

## Recommendations for Enhancing the PDTK Resource

The educators had several suggestions for strengthening the sustainability of the PDTK resource. These include (1) more training for ECE; (2) separate group lesson plan activities according to specific age groups; (3) introduction of concepts one at a time, thus breaking the lesson plan into smaller components; (4) large demonstration charts to increase the level of engagement; and (5) constant repetition of the pulse-related concepts. It was also discussed that the educator's level of interest and level of experience were key to increasing buy-in within centers. Overall, educators perceived that the PDTK has significant potential in increasing acceptance of novel food at an early age. One of the ECE's stated,



XSL•FO
RenderX

> *It's awesome. It's exactly what Saskatchewan needs right now. More people are advocating and starting young.* [ECE # 2]

### Cooks' Perceptions of the PDTK Recipes

The cooks in both CCs indicated they had prepared at least one pulse-based food every 2 weeks prior to the intervention. Neither cook saw any barrier to cooking pulses in their facility. When asked about opportunities for including more pulse-based food products, they indicated that having them on their menus and keeping cooked pulses in the freezer would increase their use in regular dishes. They indicated that it is feasible to include recipes from the PDTK resource in their menus. One of the cooks mentioned that the nutrition intervention provided more opportunities and experience in cooking new pulse recipes. They reported that most of the recipes are easy to follow and to shop for, and the food tastes good. One of the cooks indicated that she liked the lentil pizza as the children enjoyed it, and it was good for the children's health. She further mentioned that this dish can be added to their menu. The 3-bean quesadilla recipe was the favorite of one of the cooks. She said it was fun to make, whereas the other cook said that, while it was a good recipe, it was hard to cook. They both believed that the children were receptive to the recipes that they tried. Overall, they both said that they liked all the recipes, with some needing modifications.

## Discussion

### Principal Findings

This study presents findings from Phase 2 of a 3-phase project evaluating PDTK, a 12-week multicomponent nutrition intervention designed to promote consumption of pulse-based foods among preschool children in CCs. As outlined in the protocol [13], Phase 2 was conducted as a pilot study with a primary focus on assessing implementation processes and determining whether the intervention could be delivered as intended within real-world childcare settings. Accordingly, feasibility was operationalized through process-oriented outcomes, including cooks' and educators' perceptions of PDTK recipes, lesson plan evaluations, and recipe tasting activities, with particular emphasis on the practicality of recipe preparation and alignment with existing curricula. The impact of the intervention on knowledge, acceptability, and consumption of pulse-based foods was assessed. Feasibility and acceptability of the PDTK intervention for implementation in CCs were captured through lesson plan evaluations and interviews with ECEs and cooks.

To our knowledge, no recent studies have specifically examined pulse consumption in childcare settings for this age group in Canada. While the data for our study were collected a few years ago, it remains pertinent given the lack of significant interventions or policy changes that could have altered pulse consumption trends. The only notable changes are the release of the new Canada's Food Guide in 2019 and the increased promotion of plant-based diets. Evidence from a study by Hank et al [22] explored the potential impact of the 2019 Canada's Food Guide on the eating environments and food provided in early learning and CCs across Canada. This study found that representatives from CCs faced challenges in interpreting and implementing the new guidelines. As a result, no significant changes to children's dietary patterns in these settings were observed. This suggests that, despite the shift in dietary recommendations, pulse consumption in childcare environments has likely remained stable. While isolated case studies of transitions toward plant-based menus [23] have been reported in Canadian CCs, these appear to be the exception rather than the norm. Recent studies in Canada further reinforce the importance of childcare-based interventions in shaping children's dietary behaviors and daily nutrient intake [24,25].

Although some improvement was noted in the percentage of preschool children who correctly identified pulses in this study, the improvement was not noted for all pulses. This could be partly explained by the longer duration of some lessons, preschoolers' short attention span, and the small sample size. These findings contradict similar studies that introduced other food items, such as fruits and vegetables, to preschoolers [26,27]. Many of these studies indicated a significant increase in knowledge of fruit and vegetables between the pre and posttests. The differences in the findings between this pulse intervention and the fruit and vegetable interventions could be attributed to the familiarity of households with fruit and vegetables or the novelty of exposing young children to pulses.

A significantly greater amount of the regular recipes was consumed at time-point 1 than the intervention recipes; however, this difference was not seen when the consumption proportion was considered. Unlike the finding at time-point 1, the children consumed a higher quantity of the regular recipes than the intervention recipe at time-point 2, even when the consumption proportion was considered. These results contradict findings of increased consumption of lentils after repeated exposure [1]. In contrast to our study, the children in Ramsay et al [1] study were exposed to lentils 1 to 2 times per week for a total of 13 weeks. Though the children in our study were exposed to different pulse-based foods at least twice a week, the same dish was only offered every 3 weeks. In addition, Ramsay et al [1] let the children taste unseasoned lentils, while our study exposed children to a variety of pulse-based foods. These findings highlight the possibility of higher effectiveness of repeatedly exposing to one specific pulse-based dish at a time than exposure to a series of pulse-based dishes to increase intake and acceptability of pulses among young children. In addition, the pulse-based dishes introduced in this study were mixed dishes, unlike most of the studies [28,29] that reported a significant increase in acceptability and intake after repeated exposure to food. The use of mixed dishes in PDTK may have limited the positive impact of the repeated exposure on the intake and acceptability of the pulse-based dishes in our study.

The comparison between the intervention and regular recipes, which revealed lower total kilocalories and fat content in most of the intervention dishes, is not surprising given the low energy, fat, and sodium contents of pulses [30-32]. In addition, most of the regular recipes selected for comparison contain animal source foods, which could contribute to the higher fat and energy content of the regular recipes.



XSL•FO
RenderX

Unsurprisingly, some children were not willing to taste the pulse-based dishes offered to them during the sensory evaluation activities. Similar reactions have been reported before, where novel food products served to children were rejected, a phenomenon described as neophobia, which refers to the fear or avoidance of new or unfamiliar foods [33]. The children in our study were at the peak age of neophobia [34]. Food reluctance has been reported as a key challenge to offering plant-based meals in Canadian CCs [22], and food neophobia remains a well-documented barrier in early childhood nutrition, as highlighted in a recent multicountry systematic review demonstrating that neophobia contributes to children's reluctance to try unfamiliar foods and limits dietary variety [35]. The recipes with the highest rating of "Yummy," refried bean wrap and lentil smoothie, were interestingly recipes that children took part in preparing as part of the lesson activities. Similarly, though the study population was older children (5 - 7 y-old) and the finding could not solely be linked to their participation in the preparation, children rated foods they prepared "more yummy" than foods served to them [36].

Findings from ECE semistructured interviews and the lesson plan evaluations revealed that the ECEs reacted favorably to the curriculum, especially to lesson plan activities that had both sensory and hands-on components. The ECEs stated that the children particularly enjoyed preparing the snacks, role-playing, and gardening, and were highly engaged when performing these activities. The educators in this study also discussed a few factors that could potentially act as barriers to the feasibility and acceptability of the nutrition curriculum. Key barriers identified were the length of time to complete each activity and the suitability of some of the activities to meet the children's cognitive level. These findings are similar to a pilot study conducted by Sharma et al [9], a healthy nutrition and a physical activity program in 2 CCs.

In our study, the cooks from the participating CCs did not report any barriers to cooking pulses in their facility, contrary to a previous study that examined the barriers and opportunities to serving pulses in school meals, which includes time for cooking dry beans [37]. In our study, the PDTK recipes use canned pulses or dry lentils, which do not take as long as dry beans to cook. However, the need for modification to make the recipes easier to cook in childcare settings was noted in our study.

To better suit different developmental stages of preschool children, the intervention's lesson plans could be adapted by varying the complexity of activities based on cognitive and motor skills. For instance, in Lesson One, "Naming Pulses," the lesson could be tailored as follows: For younger preschoolers (ages 3 - 4 years), the lesson plan can focus on sensory exploration, where children touch and sort different pulses (eg, beans, lentils, peas) by size or color, enhancing their fine motor skills and hand-eye coordination. Teachers can also encourage language development by labeling the pulses and asking children to repeat their names, introducing basic vocabulary. For older preschoolers (ages 4 - 5 years), the activity can include problem-solving tasks such as sorting pulses by categories or counting them, along with more precise motor tasks like transferring pulses using scoops, while also fostering cognitive skills through matching games. Though our study looked at specific pulse-based dishes, we can speculate that including a wider variety of pulse-based recipes in the intervention could increase its appeal and effectiveness for children by offering more fun, diverse, and flavorful options that cater to different tastes and preferences, making the learning experience more engaging. The results were instrumental in guiding the subsequent roll-out of the intervention in phase 3 in 4 CCs. Revisions include lesson plan modifications to better align with developmental stages and age-appropriate learning, and simplified activities for younger children to emphasize sensory exploration and basic identification tasks, rather than more complex components that require higher levels of cognitive processing. These findings informed practical adjustments to the intervention design, recipe selection, and PDTK lesson plans. The lessons learned were also used to expand the program to elementary school [38]. The growing interest in plant-based dietary patterns for both health and environmental sustainability further underscores the continued relevance of interventions such as the PDTK. The growing interest in plant-based dietary patterns for both health and environmental sustainability further underscores the continued relevance of interventions such as the PDTK, particularly in light of Canadian evidence demonstrating both the feasibility of plant-based menu implementation in childcare settings and the significant role these environments play in shaping children's dietary behaviors [24,25].

## Limitations

First, due to the convenience sampling technique used and the small sample size, the generalization of the findings is limited. Second, the 2 CCs have different meal practices; in one of the centers, children are encouraged to portion their own food, whereas in the other center, food is served prior to mealtime. Additionally, children not attending all sessions during data collection or when lessons are offered may have reduced the impact of the PDTK intervention. Lack of control groups limited the study's ability to draw strong causal inferences, and future studies with a control group will better isolate the intervention's effects by comparing recipients to nonrecipients. The inclusion of cooks was limited to those working at the center, which resulted in a smaller sample size.

## Conclusions

The PDTK intervention resulted in an increase in the percentage of preschool children who correctly identified pulses, but not their intake of pulse-based foods. The findings from the pilot testing also revealed that modifications are necessary to make some of the lessons more age-appropriate. Cooks found most of the PDTK pulse-based recipes to be acceptable and were willing to include them in the centers' menus while noting that some need adjustments to make them easier for batch cooking. Further research may be necessary to determine what effects a modified PDTK intervention will have on preschool children's pulse knowledge, acceptability, and intake of the intervention pulse-based foods. In addition, a follow-up assessment can determine if behavioral changes are sustained, the durability of the intervention's effects, and whether further support is needed for long-term sustainability.



XSL•FO
RenderX

## Acknowledgments

We would like to acknowledge ECEs, children, and cooks who participated in the study. Last, we would like to thank the graduate students who helped with the study from the College of Pharmacy and Nutrition.

## Funding

This research was funded by Saskatchewan Pulse Growers.

## Conflicts of Interest

None declared.

## References

1. Ramsay SA, Roe AJ, Davis JN, Price WJ, Johnson SL. Repeated exposures and child centered nutrition phrases increases young children's consumption and liking of lentils. Food Qual Prefer 2017 Dec;62:317-322. [doi: 10.1016/j.foodqual.2017.03.002]
2. Wardle J, Herrera ML, Cooke L, Gibson EL. Modifying children's food preferences: the effects of exposure and reward on acceptance of an unfamiliar vegetable. Eur J Clin Nutr 2003 Feb;57(2):341-348. [doi: 10.1038/sj.ejcn.1601541] [Medline: 12571670]
3. Anzman-Frasca S, Savage JS, Marini ME, Fisher JO, Birch LL. Repeated exposure and associative conditioning promote preschool children's liking of vegetables. Appetite 2012 Apr;58(2):543-553. [doi: 10.1016/j.appet.2011.11.012] [Medline: 22120062]
4. Anzman SL, Rollins BY, Birch LL. Parental influence on children's early eating environments and obesity risk: implications for prevention. Int J Obes (Lond) 2010 Jul;34(7):1116-1124. [doi: 10.1038/ijo.2010.43] [Medline: 20195285]
5. Mikkelsen MV, Husby S, Skov LR, Perez-Cueto FJA. A systematic review of types of healthy eating interventions in preschools. Nutr J 2014 Jun 6;13(1):56. [doi: 10.1186/1475-2891-13-56] [Medline: 24906305]
6. Fulgoni VL, Papanikolaou Y, Fulgoni SA, Kelly RM, Rose SF. Bean consumption by children is associated with better nutrient intake and lower body weights and waist circumferences. FASEB J 2006 Mar;20(4):A621-A621. [doi: 10.1096/fasebj.20.4.A621]
7. Mudryj AN, Aukema HM, Fieldhouse P, Yu BN. Nutrient and food group intakes of Manitoba children and youth: a population-based analysis by pulse and soy consumption status. Can J Diet Pract Res 2016 Dec;77(4):189-194. [doi: 10.3148/cjdpr-2016-012] [Medline: 27744734]
8. Froehlich Chow A, Leis A, Humbert ML, Engler-Stringer R, Muhajarine N. Supporting healthy eating among rural early years children: a pulse crop pilot intervention study. J Agromedicine 2015;20(3):386-389. [doi: 10.1080/1059924X.2015.1046624] [Medline: 26237730]
9. Sharma S, Chuang RJ, Hedberg AM. Pilot-testing CATCH early childhood. Am J Health Educ 2011 Jan;42(1):12-23. [doi: 10.1080/19325037.2011.10599169]
10. Witt KE, Dunn C. Increasing fruit and vegetable consumption among preschoolers: evaluation of color me healthy. J Nutr Educ Behav 2012;44(2):107-113. [doi: 10.1016/j.jneb.2011.01.002] [Medline: 21924957]
11. Farfan-Ramirez L, Diemoz L, Gong EJ, Lagura MA. Curriculum intervention in preschool children: nutrition matters!. J Nutr Educ Behav 2011;43(4 Suppl 2):S162-S165. [doi: 10.1016/j.jneb.2011.03.007] [Medline: 21683289]
12. Horne PJ, Greenhalgh J, Erjavec M, Lowe CF, Viktor S, Whitaker CJ. Increasing pre-school children's consumption of fruit and vegetables. a modelling and rewards intervention. Appetite 2011 Apr;56(2):375-385. [doi: 10.1016/j.appet.2010.11.146]
13. Haileslassie HA, Ramikie R, Vatanparast H, et al. A nutrition intervention to promote the consumption of pulse-based foods in childcare centers: protocol for a multimethod study. JMIR Res Protoc 2020 Dec 24;9(12):e22775. [doi: 10.2196/22775] [Medline: 33361052]
14. Teresi JA, Yu X, Stewart AL, Hays RD. Guidelines for designing and evaluating feasibility pilot studies. Med Care 2022 Jan 1;60(1):95-103. [doi: 10.1097/MLR.0000000000001664] [Medline: 34812790]
15. Calfas KJ, Sallis JF, Nader PR. The development of scales to measure knowledge and preference for diet and physical activity behavior in 4- to 8-year-old children. J Dev Behav Pediatr 1991 Jun;12(3):185-190. [doi: 10.1097/00004703-199106000-00007] [Medline: 1869623]
16. Anliker JA, Laus MJ, Samonds KW, Beal VA. Parental messages and the nutrition awareness of preschool children. J Nutr Educ 1990 Jan;22(1):24-29. [doi: 10.1016/S0022-3182(12)80289-5]
17. Bélanger M, Humbert L, Vatanparast H, et al. A multilevel intervention to increase physical activity and improve healthy eating and physical literacy among young children (ages 3-5) attending early childcare centres: the healthy start-départ



XSL•FO
RenderX

santé cluster randomised controlled trial study protocol. BMC Public Health 2016 Apr 12;16(1):313. [doi: 10.1186/s12889-016-2973-5] [Medline: 27068684]
18. Birch LL, Sullivan SA. Measuring children's food preferences. J Sch Health 1991 May;61(5):212-214. [doi: 10.1111/j.1746-1561.1991.tb06015.x] [Medline: 1943045]
19. Sharma SV, Hedberg AM, Skala KA, Chuang RJ, Lewis T. Feasibility and acceptability of a gardening-based nutrition education program in preschoolers from low-income, minority populations. J Early Child Res 2015 Feb;13(1):93-110. [doi: 10.1177/1476718X14538598]
20. Ryan GW, Bernard HR. Techniques to identify themes in qualitative data. Field Methods 2003;15(1):85-109. [doi: 10.1177/1525822X02239569]
21. Braun V, Clarke V. Toward good practice in thematic analysis: avoiding common problems and be(com)ing a knowing researcher. Int J Transgend Health 2023;24(1):1-6. [doi: 10.1080/26895269.2022.2129597] [Medline: 36713144]
22. Hank I, Rossiter MD, Finch S. Influence of the 2019 Canada's food guide on the food environment in childcare and early learning centres. Can J Diet Pract Res 2023 Jun 1;84(2):93-97. [doi: 10.3148/cjdpr-2022-035] [Medline: 36866836]
23. Eckert KF, Trew V, Serediuk E, Galvez A, Haines J. Transitioning to a plant-based menu in childcare: identifying the nutritional, financial, and logistical considerations. Can J Diet Pract Res 2024 Mar 1;85(1):20-24. [doi: 10.3148/cjdpr-2023-015] [Medline: 37732547]
24. Kucab MD, Keown-Stoneman CD, Birken CS, et al. Center-based childcare in early childhood and nutritional risk in later childhood: a prospective cohort study. J Nutr 2025 Dec;155(12):4530-4539. [doi: 10.1016/j.tjnut.2025.09.028] [Medline: 41022151]
25. Bossé A MSc, Haines J PhD, RD, Duncan AM PhD, RD, Ma DWL PhD, Tugault-Lafleur CN PhD, RD. How do meals and snacks consumed in childcare contribute to children's food and nutrient intakes? Can J Diet Pract Res 2025 Sep 1;86(3):145-152. [doi: 10.3148/cjdpr-2025-008] [Medline: 40358100]
26. Sigman-Grant M, Byington TA, Lindsay AR, et al. Preschoolers can distinguish between healthy and unhealthy foods: the all 4 kids study. J Nutr Educ Behav 2014;46(2):121-127. [doi: 10.1016/j.jneb.2013.09.012] [Medline: 24246710]
27. Tande DL, Niemeier BS, Hwang JH, et al. Introducing preschool children to novel fruits and vegetables: a pilot study. J Child Nutr Manag 2014;38(1):EJ1033497 [FREE Full text]
28. Ahern SM, Caton SJ, Blundell P, Hetherington MM. The root of the problem: increasing root vegetable intake in preschool children by repeated exposure and flavour flavour learning. Appetite 2014 Sep;80:154-160. [doi: 10.1016/j.appet.2014.04.016]
29. Caton SJ, Ahern SM, Remy E, Nicklaus S, Blundell P, Hetherington MM. Repetition counts: repeated exposure increases intake of a novel vegetable in UK pre-school children compared to flavour-flavour and flavour-nutrient learning. Br J Nutr 2013 Jun;109(11):2089-2097. [doi: 10.1017/S0007114512004126] [Medline: 23110783]
30. Zhang B, Deng Z, Tang Y, et al. Effect of domestic cooking on carotenoids, tocopherols, fatty acids, phenolics, and antioxidant activities of lentils (Lens culinaris). J Agric Food Chem 2014 Dec 31;62(52):12585-12594. [doi: 10.1021/jf504181r] [Medline: 25474757]
31. Rebello CJ, Greenway FL, Finley JW. Whole grains and pulses: a comparison of the nutritional and health benefits. J Agric Food Chem 2014 Jul 23;62(29):7029-7049. [doi: 10.1021/jf500932z] [Medline: 24992700]
32. Remy E, Issanchou S, Chabanet C, Nicklaus S. Repeated exposure of infants at complementary feeding to a vegetable puree increases acceptance as effectively as flavor-flavor learning and more effectively than flavor-nutrient learning. J Nutr 2013 Jul;143(7):1194-1200. [doi: 10.3945/jn.113.175646] [Medline: 23700337]
33. Lafraire J, Rioux C, Giboreau A, Picard D. Food rejections in children: cognitive and social/environmental factors involved in food neophobia and picky/fussy eating behavior. Appetite 2016 Jan 1;96:347-357. [doi: 10.1016/j.appet.2015.09.008] [Medline: 26391004]
34. Dovey TM, Staples PA, Gibson EL, Halford JCG. Food neophobia and 'picky/fussy' eating in children: a review. Appetite 2008;50(2-3):181-193. [doi: 10.1016/j.appet.2007.09.009]
35. Del Campo C, Bouzas C, Tur JA. Risk factors and consequences of food neophobia and pickiness in children and adolescents: a systematic review. Foods 2024 Dec 30;14(1):69. [doi: 10.3390/foods14010069] [Medline: 39796359]
36. DeJesus JM, Gelman SA, Herold I, Lumeng JC. Children eat more food when they prepare it themselves. Appetite 2019 Feb;133:305-312. [doi: 10.1016/j.appet.2018.11.006]
37. Smith DK, Anne Riddle L, Kerr S, Atterberry K, Lanigan J, Miles C. Barriers and opportunities to serving pulses in school meals in Washington schools. JCN&M 2016;40(1):EJ1102969 [FREE Full text]
38. Teshome GB, Haileslassie HA, Shand P, Lin Y, Lieffers JRL, Henry C. Pulse-based nutrition education intervention among high school students to enhance knowledge, attitudes, and practices: pilot for a formative survey study. JMIR Form Res 2023 May 31;7:e45908. [doi: 10.2196/45908] [Medline: 37256666]

## Abbreviations

**CC:** childcare center
**ECE:** early childhood educator
**GEE:** generalized estimating equation



XSL•FO
RenderX

**PDTK:** Pulse Discovery Toolkit

*Edited by M Balcarras; submitted 27.Dec.2023; peer-reviewed by A Hassan, G Wangge, K Shillington, S Mitra; revised version received 29.Mar.2026; accepted 30.Mar.2026; published 22.May.2026.*

*Please cite as:*
*Haileslassie HA, Ramikie R, Vatanparast H, Ramdath DD, Froehlich Chow A, Shand P, Engler-Stringer R, Lieffers JRL, Hood-Niefer S, Henry C*
*Pulse Discovery Toolkit, a Multicomponent Nutrition Intervention for Preschool Children in Childcare Centers: Mixed Methods Pilot Study*
*Interact J Med Res 2026;15:e55866*
*URL: https://www.i-jmr.org/2026/1/e55866*
*doi:10.2196/55866*

© Hiwot Abebe Haileslassie, Renee Ramikie, Hassan Vatanparast, D Dan Ramdath, Amanda Froehlich Chow, Phyllis Shand, Rachel Engler-Stringer, Jessica R L Lieffers, Shannon Hood-Niefer, Carol Henry. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 22.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Prevalence and Associated Factors of Excessive Dietary Supplement Use Among Japanese Adults: Cross-Sectional Study

Minami Sugimoto[1], MPH, PhD; Keiko Asakura[2], MD, MPH, PhD; Nana Shinozaki[3], MPH, RD, PhD; Kentaro Murakami[3], BEd, MSc, PhD

[1]Department of Environmental and Occupational Health, School of Medicine, Toho University, Tokyo, Japan

[2]Department of Preventive Medicine, School of Medicine, Toho University, Omori-Nishi 5-21-16, Ota-ku, Tokyo, Japan

[3]Department of Social and Preventive Epidemiology, School of Public Health, The University of Tokyo, Tokyo, Japan

**Corresponding Author:**
Keiko Asakura, MD, MPH, PhD
Department of Preventive Medicine, School of Medicine, Toho University, Omori-Nishi 5-21-16, Ota-ku, Tokyo, Japan

## Abstract

**Background:** The use of dietary supplements (DSs) can lead to an excessive intake of certain nutrients, posing potential health risks. However, studies are scarce on whether DS users adhere to recommended doses provided by the manufacturer and what factors are associated with overconsumption of DS.

**Objective:** By leveraging purchase history data to estimate DS use precisely, this study aimed to investigate the factors associated with DS consumption exceeding manufacturer-recommended doses and examine the prevalence of excess nutrient intake among DS users.

**Methods:** An online survey was conducted from November to December 2024 among 2002 adults (aged 18 - 74 years) with a history of buying one of the 25 major DS products and who had used it in the previous month or regularly. Self-reported daily DS consumption was assessed using a questionnaire and compared to the manufacturer-recommended doses indicated in the package. Using multivariate logistic regression analysis, the associations between DS consumption exceeding the manufacturer-recommended doses and sociodemographic factors were examined.

**Results:** The prevalence of DS users exceeding tolerable upper intake levels (ULs), defined according to the Dietary Reference Intakes for Japanese, was calculated based on vitamin and mineral intake from DS only. Consequently, 371 (18.5%) of the 2002 participants consumed DS above the manufacturer-recommended dose. Consumption of DS above the recommended dose was associated with middle-aged, part-time or full-time employment, use of tablet-form DS, especially single water-soluble vitamin tablets, use of DS for 6 months or more, and intentional consumption above the recommended dose. For 1705 individuals consuming DS with UL nutrients, 17.4% (n=297) exceeded the manufacturer-recommended dose. Of these, 61.9% (184/297) surpassed UL by ≥1 nutrient.

**Conclusions:** Middle-aged, part-time or full-time jobs, using tablet-form DS, and using DS for 6 months or more were associated with DS consumption exceeding the manufacturer-recommended dose, which may lead to excessive nutrient intake.

***(Interact J Med Res 2026;15:e82623)*** doi:10.2196/82623

**KEYWORDS**

dietary supplement; Japanese; risk factors; nutrients; dietary reference intake

## Introduction

Over the last few decades, the use of dietary supplements (DSs) has been increasing in many countries [1-4]. For example, the proportion of DS users in the United States increased from 51.8% in 2011 - 2012 to 61.4% in 2021 - 2023 [3], while in Australia, it increased from 28.5% in 2011 - 2012 to 33.6% in 2023 [4]. Although DSs are widely used to promote health and prevent nutrient deficiencies, concerns have emerged regarding their safety due to DS overuse and the risk of excessive nutrient intake [5,6]. Since primary sources of vitamins and minerals are a normal diet [7], having nutrients from DS in addition to diet might increase the risk of exceeding the tolerable upper intake level (UL). Excessive nutrient intake could bring adverse health effects such as liver dysfunction by vitamin A and hypercalcemia by vitamin D and calcium [8]. Therefore, understanding the behaviors and factors that contribute to DS overuse and excessive nutrient intake is a pressing public health issue.

Previous studies have investigated the prevalence [6,7,9-15] and risk factors [16] of exceeding the ULs when considering DS consumption. For instance, approximately 10% of DS users surpass the UL for iron or zinc, and <5% exceed the UL for calcium or magnesium among adults in the United States [9].



XSL•FO
RenderX

In Ireland, the risk of excess was most notable for iron and vitamin $B_6$ among women taking DS [12]. Although high-dose DS (ie, containing higher content of nutrients in one tablet or capsule) itself constitutes a risk for excessive nutrient intake [17], one possible cause of excessive nutrient intake among DS users is the overconsumption of DS products. In the United States and European countries, regulations mandate that DS product labels display the manufacturer-recommended dose (written as "serving size" or "recommended daily intake") to safeguard consumers [18,19]. However, studies on whether DS users adhere to manufacturer-recommended doses are scarce. For example, a 2012 survey by the Consumer Commission, Cabinet Office, Japan, reported that 2.4% of 8064 DS users consumed DS exceeding doses recommended by the manufacturer [20]. The prevalence of adherence to manufacturer-recommended doses may have been underestimated because the survey relied on self-reports to assess adherence to DS use, namely asking participants whether their consumption aligned with, below, or above the recommended dose.

Most previous studies [2-4,7,9-16], as well as the aforementioned Japanese survey [20], relied on self-reported data to collect information on the consumption of DS products and their use. This approach has several limitations. Participants may misreport or forget the details regarding DS use, making it challenging to identify the exact products used [21]. Furthermore, DS databases have rapidly become outdated because of the rapid turnover of products in the market [21]. These methodological weaknesses hinder the accurate estimation of DS consumption and compliance. To address these issues, purchase history data can be useful for precisely identifying DS products. Using purchase history data enables us to ask about the usage habits of specific products among participants without self-reporting the product names. This prevents the generation of data that is unanalyzed due to unknown product names. By leveraging purchase history data, this study aimed to address these gaps, with the following objectives: (1) to investigate the factors associated with DS consumption exceeding the manufacturer-recommended doses indicated on the product label or manufacturer website and (2) to examine the prevalence of nutrient intake exceeding the tolerable upper limits among DS users. This study provides a methodologically robust alternative to traditional self-reporting approaches. Since no previous study has investigated factors associated with overuse of DS, we analyzed, as a first step, variables related to sociodemographic and socioeconomic status and dietary supplement use habits.

# Methods

## Participants and Overview of Study Design

This cross-sectional study used data obtained through a web-based questionnaire survey. Data were collected between November 14 and December 1, 2024. This study was prepared according to the Checklist for Reporting Results of Internet e-Surveys (CHERRIES) [22]. The process of participant selection included identifying commonly used DS and subsequently selecting participants who reported using them. The survey procedure is detailed in the following sections. Target participants were Japanese adults (≥18 years) who consumed at least one DS product in the previous month. The target sample size was set at 2000 adults. This sample size allowed for reasonably precise estimates of the prevalence of nutrient intake exceeding the ULs, while also remaining feasible within financial constraints.

Survey participants were recruited from a registered consumer panel. The panelists routinely recorded their purchased products (such as food, daily necessities, and DSs) through a dedicated smartphone app. This consumer panel comprised primary shoppers in each household, with 70% being females and 75% being in their 20s to 40s. The panelists were distributed across Japan, closely reflecting the national population distribution.

Although the consumer panel was managed by Research and Innovation Co. Ltd., the data collection process (including participant invitation, survey distribution, and data management) was conducted using the research infrastructure of INTAGE Inc., a major market research company in Japan and the parent company of Research and Innovation Co. Ltd. This was because Research and Innovation Co. Ltd. did not have a system for asking users detailed questions, whereas INTAGE Inc. did.

## Ethical Considerations

This study was approved by the ethics committee of the Toho University Faculty of Medicine (approval no. A24059; approved on October 7, 2024). Informed consent was obtained from all the participants on the website. To ensure privacy and confidentiality, the research team received only anonymized data from the survey company (INTAGE Inc.), and no personally identifiable information was accessible to the researchers. All data were handled in accordance with institutional ethical guidelines. The participants were compensated with standard incentives from Research and Innovation Co. Ltd, after completing the survey.

## Survey Procedure

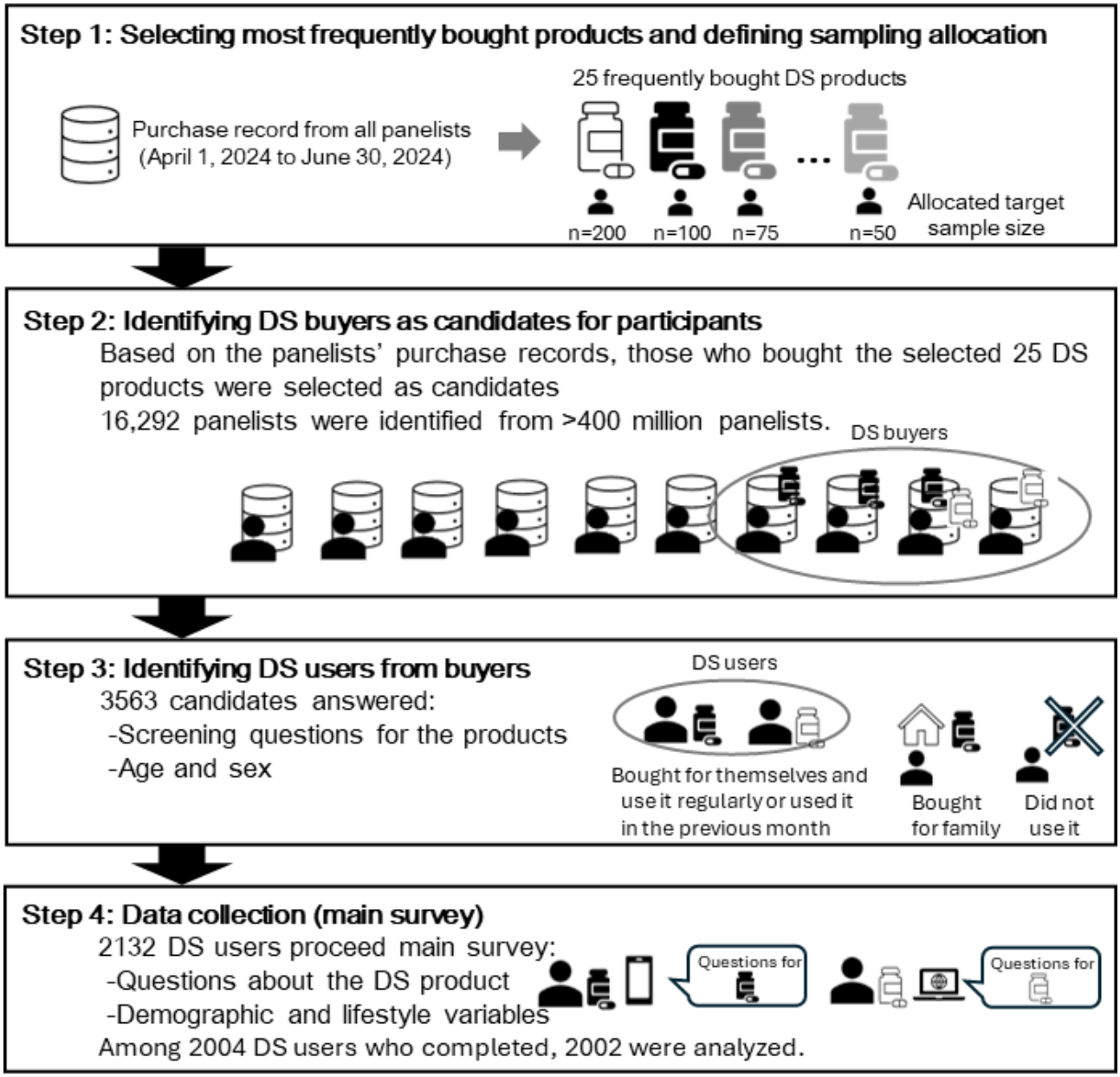
### *Overview*

Data were collected in 4 steps (Figure 1). The first and second authors (MS and KA) prepared the questionnaires. Each participant reported information about one of the selected DS products.



XSL•FO
RenderX

**Figure 1.** Survey procedure. DS: dietary supplement.




### *Step 1: Selecting the Most Frequently Bought Products and Defining Sampling Allocation*

Due to the lack of comprehensive and nationwide statistical data on the consumption of DS products, the most frequently purchased DS products are unknown. Thus, using purchase record data from all panelists for the past 3 months (April 1, 2024, to June 30, 2024), the second author (KA) identified the most frequently purchased DS products across the following 5 categories: liquid form supplements, multivitamin tablets, water-soluble vitamin tablets, fat-soluble vitamin tablets, and mineral tablets. The anticipated number of respondents was first allocated to each of the 5 categories to ensure accurate estimation of the prevalence of excessive intake. Considering the actual sales volumes, the anticipated number of respondents was allocated as follows: 300 for liquid-form DSs, 400 for multivitamin tablets, 600 for water-soluble vitamin tablets, 200 for fat-soluble vitamin tablets, and 500 for mineral tablets. To meet the target sample sizes, products were selected in descending order of sales volume within each category. The number of products selected in each category was as follows: 2 liquid-form DS, 6 multivitamin tablets, 6 water-soluble vitamin tablets, 7 fat-soluble vitamin tablets, and 4 mineral tablets, totaling 25 products. The names of the selected 25 DS products and the expected number of respondents for each product are presented in Table S1 in Multimedia Appendix 1.

### *Step 2: Identifying DS Buyers as Candidates for Participants*

The buyers of the 25 DS products selected in step 1 were identified from over 4 million panelists by INTAGE Inc as candidates for participants. Panelists were selected as buyers of DS if they had at least one purchase history of ≥1 DS among 25 selected DS products in the past 3 months. To avoid duplicate responses from individuals, those with a record of purchasing 2 or more selected DS products were randomly assigned as potential respondents to one of the selected DS products. INTAGE Inc. operates a random allocation. Consequently,



XSL•FO
RenderX

16,292 panelists were identified as DS buyers (ie, candidates for participation).

### *Step 3: Identifying DS Users From Buyers (Screening)*

INTAGE Inc. sent invitation emails to candidates. Participants completed the questionnaire via a survey platform provided by INTAGE Inc using their own smartphone, tablet, or PC. Through identity authentication, access to the questionnaire was restricted to the individual panelist surveyed. During the survey, the collected data were encrypted and stored in a database at INTAGE Inc. To screen DS users, the candidates were asked the following questions after a question about their sex and age:

- Question 1: We would like to ask you about the dietary supplement [product name] you have purchased in the past 3 months. Did you purchase this product for your use? (options: yes, for myself; no, for my family)
- Question 2: Do you still use this product regularly? (options: yes or no)
- Question 3: Have you taken this product in the last month? (options: yes or no)

Those who answered "yes" to question 1 and also responded "yes" to either or both questions 2 and 3 were identified as DS users. Among the 16,292 candidates, 3563 (21.9%) panelists answered the screening questions. After screening the questions, 1431 panelists were excluded, and 2132 (13.1%) of the 16,292 panelists were identified as DS users and proceeded to the main survey.

### *Step 4: Data Collection (Main Survey)*

A total of 2132 participants were asked about the use of one of the 25 selected DS products. Among them, 106 participants did not complete the questionnaire, and 22 panelists were excluded because of unreliable data and an extremely short response time. Overall, 2004 participants completed the main survey at a completion rate of 94% (2004/2132). The main survey comprised 50 questions. The number of items per page was set to one to increase readability on the screen. Thus, the survey consisted of 50 screens in total.

## Assessment of DS Use

The participants were asked to respond to a series of questions regarding one of the 25 DS products purchased. To simplify the survey procedure, we did not ask questions about the use of other products. To help participants accurately identify the product in question, the survey screen displayed an image of the product package for each set of questions and referred to a specific product name in the text. If the product package shows the manufacturer-recommended dose, this information was masked in the image provided.

To estimate DS consumption per day, they were first asked for their consumption at one time with the following question, with the appropriate unit of measurement (such as tablet or milliliter) automatically displayed depending on the product:

- We would like to ask you about the dietary supplement [product name] you have purchased in the past 3 months. Please answer the subsequent questions without looking at the product container or packaging. How much do you consume at one time? If you take it multiple times a day, please report the amount for a single intake, not the daily total.

The participants entered the quantity as an integer. Next, they were asked about their consumption frequency: "How frequently do you consume [product name]?" The frequency was selected from the following predefined 12 categories, from "more than six times per day" to "less than once per month," as shown below. The reference period (ie, consumption in the last month) was not specified in these questions because we had already asked the participants whether they had used the DS in the previous month and whether they had used it regularly in the screening.

DS consumption per day was calculated by multiplying the number of tablets or milliliters of liquid consumed at a time by the consumption frequency. The multiplied numbers according to the consumption frequency were as follows: more than six times per day=6, five times per day=5, four times per day=4, three times per day=3, two times per day=2, once per day=1, four to six times per week=5/7, two or three times per week=2.5/7, once per week=1/7, two or three times per month=2.5/30, once per month=1/30, and less than once per month=0.5/30.

Furthermore, the duration of use, checking the manufacturer-recommended dose before use, and self-perception of their DS consumption were asked via the following questions, respectively:

- For [product's name], how long have you been using it for? (options: <3 months, 3 months to <6 months, 6 months to <1 year, or ≥1 year)
- For [product's name], before you started using the product, did you check the manufacturer-recommended doses of the product? (options: yes or no)
- For [product name], compared to the manufacturer-recommended doses, your consumption is… (options: more, same, or less than the manufacturer-recommended doses, unsure, or no manufacturer-recommended doses provided)

## The Manufacturer-Recommended Doses of Products

All 25 selected products indicated the manufacturer-recommended dose, although labeling of the manufacturer-recommended dose is not mandated for DS in Japan, with some exceptions [23]. Among the 25 selected products, 18 are labeled as "food with nutrient function claims (*Eiyo-Kino-Shokuhin* in Japanese)," which are required to display a manufacturer-recommended dose [24]. The other 2 products are approved as "designated quasi-drugs" by the Ministry of Health, Labour, and Welfare and are required to indicate the approved dosage and administration on the label [25]. They can be bought at a drugstore, grocery store, or convenience store without a prescription and are usually used as DS. The other 5 products were voluntarily shown the manufacturer-recommended doses. Table S1 in Multimedia Appendix 1 shows which products were labeled as "food with nutrient function claim," "designated quasi-drugs," or neither. The manufacturer's website was searched using the product's



XSL•FO
RenderX

name and package image. The names and package images used in the survey were provided to researchers by INTAGE Inc. When searching the manufacturers' websites, we ensured that the packaging images on the sites matched the packaging images of the products used in the survey. The manufacturer-recommended doses of each product were extracted from the website of the first author (MS). Although evaluating the appropriateness of the manufacturer-recommended doses was outside the scope of this study, no product exceeded the UL for any nutrient at the manufacturer-recommended doses.

### Estimation of Nutrient Intake From DSs

Nutrient intake was calculated for the following 16 nutrients using the UL established in the Dietary Reference Intakes (DRIs) for Japanese 2025 [8]: vitamins A, D, E, $B_6$, niacin, folic acid, calcium, magnesium (for intakes derived from DS only), phosphorus, zinc, copper, manganese, iodine, selenium, chromium, and molybdenum. In Japanese DRIs, UL was defined based on "lowest observed adverse effect level" or "no observed adverse effect level" and uncertainty factor by the committee for DRIs composed of researchers regarding nutrition, medicine, and public health [8]. Nutrient intake from the DS was estimated by multiplying the daily DS consumption (number of tablets or mL) by the nutrient content of the products (per tablet or mL). The contents of these vitamins and minerals per tablet or milliliter in each product were obtained from the product's website. There were no products providing phosphorus content. Thus, intakes of 15 nutrients were calculated.

### Assessment of Demographic and Lifestyle Variables

Sex and age were assessed during the screening. Other information regarding demographic and lifestyle factors was provided in the main survey. The sex at birth was selected as either male or female. Age, body weight, and body height were determined using numerical values and integers. When filling in their body weight and height, the participants were instructed to round to the nearest whole number. To prevent participants from discontinuing the survey due to their reluctance to disclose their body weight or height, these two items were not designated as mandatory responses. The following information was also collected by the questionnaire (categorization shown in parentheses): education level (junior high school, senior high school, 2-year college or technical professional school, university or higher, others, refusal to answer), employment status (unemployed, student, part-time job, full-time job), medical history (hypertension, stroke or cerebral hemorrhage, dyslipidemia, diabetes mellitus, hyperuricemia/gout, fracture or osteoporosis, liver disease, kidney disease, lung disease, heart disease, gastrointestinal disease, cancer, other, refusal to answer; multiple choice allowed); smoking status (never, past, current [<20 cigarettes per day], current [≥20 cigarettes]), drinking habit (not at all, hardly, sometimes, every day), pregnant (yes, no, refusal to answer), breastfeeding (yes, no, refusal to answer). In Japan, alcohol consumption and smoking are legally permitted from the age of 20. Therefore, following INTAGE Inc.'s policy, questions regarding drinking and smoking habits were not asked of participants aged <20 years (n=3).

Age was then categorized into 4 groups (18 - 34, 35 - 49, 50 - 64, and ≥65 years). The ≥65-year group was classified separately to reflect the older adult population, while those aged 18 - 64 years were categorized into 3 groups, approximately 15-year intervals, to capture distinct developmental and lifestyle stages. BMI ($kg/m^2$) was calculated as body weight (kg) divided by the square of body height (m). Weight status was categorized as underweight (<18.5 $kg/m^2$), normal (18.5 - 24.9 $kg/m^2$), overweight or obese (≥25 $kg/m^2$) [8], or missing (who did not answer body weight, height, or both). Educational level was categorized into 4 groups: junior or senior high school, 2-year college or technical school, university or higher, and other or refusal to answer. Medical history was categorized into 3 categories: yes (having any medical history), no, and refusal to answer. Pregnancy and breastfeeding were categorized into 2 groups: yes and no/refusal. The 3 participants aged <20 years were classified as "never smokers" for smoking habits and "not at all" for alcohol consumption.

Food literacy associated with adherence to the manufacturer-recommended regimen was assessed using the Japanese version of the Self-Perceived Food Literacy Scale [26]. The Japanese version was developed and used in a previous study [27]. This scale is an expert-based, theory-driven, and validated tool for measuring food literacy in healthy eating that focuses on 8 domains. Participants were asked to respond on a 5-point Likert scale (1=not at all or never; 5=yes or always). The food literacy score was calculated as the average of all 29 items. Negative items were reversed in the calculation. Therefore, the higher the score, the higher the food literacy. Possible scores ranged from 1 to 5. Participants were categorized into 2 groups based on their scores: those below the median and those at or above the median.

### Analyzed Participants

Participants were excluded from the analysis if their amount of DS consumption was extremely high due to the possibility of incorrect responses, specifically those who reported taking a 60- or 90-day package more than 6 times or more per day (ie, ≥360 or ≥540 tablets per day) (n=2). Daily consumption of DS compared to the recommended dose was calculated as the dietary consumption of DS (ie, the number of tablets or mL of liquid per day) divided by the manufacturer-recommended daily dose of the product. When the daily consumption of DS compared to the recommended dose (%) was over 100, the participants were categorized into the "exceeding–manufacturer-recommended-dose" group (who consumed DS exceeding the recommended dose). Others were classified into the "within–manufacturer-recommended-dose" group (who consumed DS at the same or below the recommended dose).

### Statistical Analysis

All statistical analyses were performed using the SAS statistical software (version 9.4; SAS Institute Inc). Significant results were defined as 2-tailed *P* values <.05. Variables of demographic and lifestyle factors and food literacy scores are presented as frequencies and percentages for categorical variables and means and SDs for continuous variables. First,



XSL•FO
RenderX

these variables were compared between 2 groups: the within–manufacturer-recommended-dose group and exceeding–manufacturer-recommended-dose group. Statistical differences were analyzed using the Student *t* test for continuous variables and the chi-square test for categorical variables.

Univariate and multivariate logistic regression models were used to evaluate the factors associated with DS consumption exceeding manufacturer-recommended doses. Individuals in the manufacturer-recommended-dose group served as the reference group. The explanatory variables were demographic and lifestyle variables, namely, age, sex, weight status, educational level, employment status, medical history, smoking status, and drinking habits. In addition, the following variables were examined as explanatory variables: form of DS, featured nutrients of DS tablets, duration of use, checking of the manufacturer-recommended dose before use, self-perception of DS consumption, and food literacy score. Featured nutrients of tablet form DS were categorized as follows: "multivitamin/multimineral" containing 3 or more vitamins or minerals; "folic acid and iron" containing folic acid and iron as the primary nutrients; "single fat-soluble vitamins" containing <3 fat-soluble vitamins as the primary nutrients; "single water-soluble vitamins" containing <3 water-soluble vitamins as the primary nutrients; and "single mineral" containing <3 minerals as the primary nutrients.

In the multivariate model, the following covariates were mutually adjusted for: sex, age (18 - 34, 34 - 49, 50 - 64, or ≥65 years), weight status (<18.5 kg/m$^2$, 18.5 - 24.9 kg/m$^2$, ≥25 kg/m$^2$, or missing), educational level (junior or senior high school, 2-year college or technical professional school, university or higher, or others or refusal to answer), employment status (unemployed, student, part-time job, or full-time job), medical history (yes, no, or refusal to answer), smoking status (never, past, current [<20 cigarettes a day], or current [≥20 cigarettes a day]), drinking habit (not at all, hardly, sometimes, or every day), and form of DS (liquid or tablet). The form of the DS was adjusted for differences in user characteristics, including consumption compared to the manufacturer-recommended doses, sex, weight status, and whether users checked the manufacturer-recommended doses before use (Table S2 in Multimedia Appendix 1). When one of the main variables, such as the duration of DS use, checking the manufacturer-recommended dose before use, and self-perception of DS consumption compared with the manufacturer-recommended dose, or food literacy score, was included in the model, it was analyzed alongside 9 covariates. When the featured nutrients of tablet form DS were the main variables, the form of DS was not adjusted because this analysis only included participants who consumed tablet-type DS. Additionally, linear trends of odds ratios were tested for age, educational level, employment status, smoking status, drinking habit, and duration of DS use. In the regression analysis, ordinal variables (ie, 1, 2, 3, and 4) were used as explanatory variables for educational level, employment status, smoking status, and drinking habit. The median value of each category was used for age and duration of use.

Nutrient intake from DS was described as the median and 25th and 75th percentiles for those with ULs. The medians and 25th and 75th percentiles were calculated for those who consumed DS containing each nutrient with ULs. For example, vitamin A intake was described in participants who consume DS containing vitamin A. In other words, participants who consumed DS without vitamin A were excluded from the calculation of vitamin A intake. In addition, the number and proportion of participants who exceeded the UL were calculated for all participants and within– or exceeding–manufacturer-recommended-dose group.

For sensitivity analysis, the same analysis was conducted after excluding participants who reported consuming DS five times or more than the manufacturer-recommended dose. Overall, 1821 participants were analyzed after excluding those who consumed DS five times or more than the manufacturer-recommended dose (n=183).

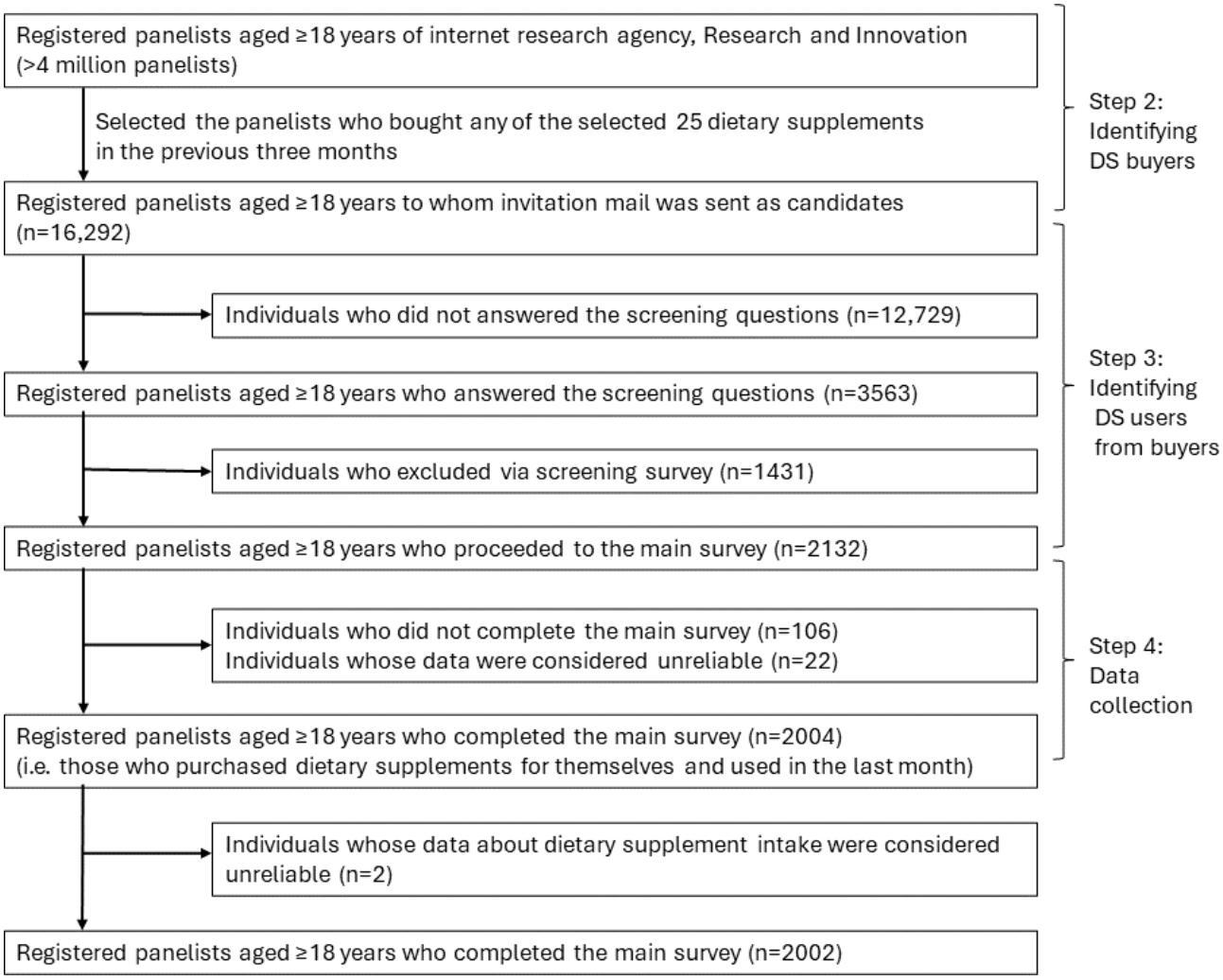
## Results

A flow diagram of the participant selection process is shown in Figure 2. The mean age was 43.7 (SD 10.7) years, and half of the participants were aged 34 - 49 years (Table 1). The mean BMI was 21.6 (SD 3.6) kg/m$^2$ (n=1755, for those who answered both body weight and height); 59.5% (1192/2002) of the participants were categorized as having normal weight, while 12.3% (247/2002) did not answer their body weight, height, or both. The participants were predominantly female (1514/2002, 75.6%). The mean age and BMI were not significantly different between the within– and exceeding–manufacturer-recommended-dose groups (43.6 [SD 10.6] vs 44.2 [SD 11.1] years; 21.6 [SD 3.6] vs 21.8 [SD 3.4] kg/m$^2$, respectively). The exceeding–manufacturer-recommended-dose group had a higher proportion of participants aged 51 - 64 years and with drinking habits "every day." There were no statistically significant differences in other demographic and lifestyle variables between the within– and exceeding–manufacturer-recommended-dose groups.



XSL•FO
RenderX

**Figure 2.** Flowchart of study participants' selection. In step 2, those with a record of purchasing 2 or more selected DS products were randomly assigned as potential respondents to only one of the selected DS products. DS: dietary supplement.






XSL•FO
RenderX

**Table .** Characteristics of dietary supplement users (n=2002) according to dietary supplement consumption compared to the recommended dose[a].

| Variables | All (n=2002), n (%) | Within–manufacturer-recommended dose[b] (n=1631), n (%) | Exceeding–manufacturer-recommended dose[c] (n=371), n (%) | *P* value[d] |
|---|---|---|---|---|
| Age (years) | | | | .04 |
| 18 - 34 | 421 (21.0) | 337 (20.7) | 84 (22.6) | |
| 34 - 49 | 956 (47.8) | 803 (49.2) | 153 (41.2) | |
| 50 - 64 | 567 (28.3) | 444 (27.2) | 123 (33.2) | |
| ≥65 | 58 (2.9) | 47 (2.9) | 11 (3.0) | |
| Weight status | | | | .35 |
| Underweight (BMI<18.5 kg/m$^2$) | 298 (14.9) | 253 (15.5) | 45 (12.1) | |
| Normal (BMI 18.5 - 24.9 kg/m$^2$) | 1192 (59.5) | 961 (58.9) | 231 (62.3) | |
| Overweight or obesity (BMI≥25 kg/m$^2$) | 265 (13.2) | 213 (13.1) | 52 (14.0) | |
| Missing | 247 (12.3) | 204 (12.5) | 43 (11.6) | |
| Sex | | | | .03 |
| Male | 488 (24.4) | 381 (23.4) | 107 (28.8) | |
| Female | 1514 (75.6) | 1250 (76.6) | 264 (71.2) | |
| Educational level | | | | .27 |
| Junior or senior high school | 670 (33.5) | 539 (33.0) | 131 (35.3) | |
| 2-year college or technical school | 588 (29.4) | 472 (28.9) | 116 (31.3) | |
| University or higher | 669 (33.4) | 561 (34.4) | 108 (29.1) | |
| Others or refusal to answer | 75 (3.7) | 59 (3.6) | 16 (4.3) | |
| Employment status | | | | .05 |
| Unemployed | 496 (24.8) | 424 (26.0) | 72 (19.4) | |
| Student | 27 (1.3) | 20 (1.2) | 7 (1.9) | |
| Part-time job | 533 (26.6) | 427 (26.2) | 106 (28.6) | |
| Full-time job | 946 (47.3) | 760 (46.6) | 186 (50.1) | |
| Medical history | | | | .59 |
| No | 1276 (63.7) | 1034 (63.4) | 242 (65.2) | |
| Yes | 629 (31.4) | 520 (31.9) | 109 (29.4) | |
| Refusal to answer | 97 (4.8) | 77 (4.7) | 20 (5.4) | |
| Smoking status | | | | .05 |
| Never | 1358 (67.8) | 1125 (69.0) | 233 (62.8) | |
| Past | 369 (18.4) | 297 (18.2) | 72 (19.4) | |
| Current (<20 cigarettes a day) | 219 (10.9) | 168 (10.3) | 51 (13.7) | |
| Current (≥20 cigarettes a day) | 56 (2.8) | 41 (2.5) | 15 (4.0) | |
| Drinking habit | | | | .002 |
| Not at all | 647 (32.3) | 524 (32.1) | 123 (33.2) | |
| Hardly | 478 (23.9) | 411 (25.2) | 67 (18.1) | |



XSL•FO
RenderX

| Variables | All (n=2002), n (%) | Within–manufacturer-recommended dose[b] (n=1631), n (%) | Exceeding–manufacturer-recommended dose[c] (n=371), n (%) | *P* value[d] |
|---|---|---|---|---|
| Sometimes | 616 (30.8) | 501 (30.7) | 115 (31.0) | |
| Every day | 261 (13.0) | 195 (12.0) | 66 (17.8) | |
| Pregnant (n, % of females) | | | | .89 |
| Yes | 42 (2.8) | 35 (2.8) | 7 (2.7) | |
| No or refusal to answer | 1467 (97.2) | 1215 (97.2) | 257 (97.3) | |
| Breastfeeding (n, % of females) | | | | .16 |
| Yes | 48 (3.2) | 36 (2.9) | 12 (4.5) | |
| No or refusal to answer | 1461 (96.8) | 1214 (97.1) | 252 (95.5) | |
| Form of dietary supplement | | | | <.001 |
| Liquid | 301 (15.0) | 287 (17.5) | 14 (3.8) | |
| Tablet | 1702 (85.0) | 1345 (82.5) | 357 (96.2) | |
| Featured nutrients of the dietary supplement[e] (tablet form only) | | | | <.001 |
| Multivitamins/multimineral | 486 (24.3) | 394 (24.2) | 92 (24.8) | |
| Folic acid and iron | 129 (6.4) | 108 (6.6) | 21 (5.7) | |
| Single fat-soluble vitamins | 169 (8.4) | 132 (8.1) | 37 (10.0) | |
| Single water-soluble vitamins | 470 (23.5) | 347 (21.3) | 123 (33.2) | |
| Single minerals | 447 (22.3) | 363 (22.3) | 84 (22.6) | |
| Duration of the dietary supplement use | | | | <.001 |
| <3 months | 484 (24.2) | 420 (25.8) | 64 (17.3) | |
| 3 months to <6 months | 289 (14.4) | 245 (15.0) | 44 (11.9) | |
| 6 months to <1 year | 375 (18.7) | 303 (18.6) | 72 (19.4) | |
| ≥1 year | 855 (42.7) | 664 (40.7) | 191 (51.5) | |
| Checking the manufacturer-recommended dose before use: “Before you started using the product, did you check the recommended dose of the product?” | | | | <.001 |
| Yes | 1686 (84.2) | 1351 (82.8) | 335 (90.3) | |
| No | 316 (15.8) | 280 (17.2) | 36 (9.7) | |
| Self-perception of their dietary supplement consumption: “Compared to the recommended dose, your consumption is...” | | | | <.001 |
| More than the recommended dose | 82 (4.1) | 25 (1.5) | 57 (15.4) | |
| Same as the recommended dose | 1268 (63.3) | 1056 (64.7) | 212 (57.1) | |
| Less than the recommended dose | 240 (12.0) | 200 (12.3) | 40 (10.8) | |



XSL•FO
RenderX

| Variables | All (n=2002), n (%) | Within–manufacturer-recommended dose[b] (n=1631), n (%) | Exceeding–manufacturer-recommended dose[c] (n=371), n (%) | *P* value[d] |
|---|---|---|---|---|
| Unsure | 387 (19.3) | 326 (20.0) | 61 (16.4) | |
| No recommended dose provided | 25 (1.2) | 24 (1.5) | 1 (0.3) | |
| Food literacy score[f] | | | | .49 |
| Less than median score | 1020 (50.9) | 837 (51.3) | 183 (49.3) | |
| Same as the median score or more | 982 (49.1) | 794 (48.7) | 188 (50.7) | |

[a]Participants were panelists who bought one of the 25 selected dietary supplements in the previous 3 months and used it within the past month or regularly. Information was collected from one of the 25 dietary supplements used by each participant.

[b]Those whose daily intake of the dietary supplement was equal to or below the manufacturer-recommended dose.

[c]Those whose intake exceeded the manufacturer-recommended dose.

[d]*P* values by chi-square test for categorical variables. Statistical significance was set at $P<.05$.

[e]"Multivitamins" contained 3 or more vitamins or minerals. "Folic acid and iron" contained folic acid and iron as the primary nutrients. "Single fat-soluble vitamins" contained <3 fat-soluble vitamins as the primary nutrients. "Single water-soluble vitamins" contained <3 water-soluble vitamins as the primary nutrients. "Single mineral" contained <3 minerals as the primary nutrients.

[f]The median score among the participants was 3.17, with possible scores ranging from 1 to 5.

Among 2002 participants, 18.5% (371/2002) exceeded the manufacturer-recommended dose. Regarding the variables related to DS use, the exceeding–manufacturer-recommended-dose group had a higher proportion of individuals who used tablet-form DS, especially single water-soluble vitamin products, and those who used the products for ≥6 months. In addition, the proportion of participants who checked the manufacturer-recommended doses was higher in the exceeding–manufacturer-recommended-dose DS group than in the within–manufacturer-recommended-dose group (90.3% [335/371] vs 82.8% [1351/1631]). The proportion of participants who thought their consumption was "more than the manufacturer-recommended dose" was higher in the exceeding–manufacturer-recommended group (57/371, 15.4%) than in the within–manufacturer-recommended group (25/1631, 1.5%). Moreover, 57.1% (212/371) and 10.8% (40/371) of those in the exceeding–manufacturer-recommended-dose group thought that their consumption was the same or below the recommended dose, respectively. In the exceeding–manufacturer-recommended-dose group, the daily DS consumption compared with the manufacturer-recommended dose (%) ranged from 394% for single water-soluble vitamin tablets to 543% for folic acid and iron tablet DS (Table S3 in Multimedia Appendix 1).

The unadjusted and adjusted odds ratios (AORs) for exceeding the manufacturer-recommended dose are presented in Table 2. From the multivariate logistic regression analysis, participants aged 50 - 64 years (AOR 1.45, 95% CI 1.10-1.92), those with part-time jobs (AOR 1.58, 95% CI 1.13-2.21), or those with full-time jobs (AOR 1.49, 95% CI 1.09-2.03) were likely to consume DS exceeding the manufacturer-recommended doses compared with those aged 18 - 34 years and those who were unemployed. In addition, participants who used a tablet form DS (AOR 6.02, 95% CI 3.46-10.47), especially a single water-soluble vitamin tablet DS (AOR 1.51, 95% CI 1.11-2.06), were likely to exceed the manufacturer-recommended doses compared to those using liquid-form DS and multivitamin/multimineral tablets. Moreover, using DS for 6 months to <1 year or ≥1 year, and those who thought they took "more than" the manufacturer-recommended doses were associated with exceeding the manufacturer-recommended doses.



XSL•FO
RenderX

**Table .** Factors associated with dietary supplement consumption exceeding the manufacturer-recommended doses among 2002 dietary supplement users[a].

| Variables | Crude model[b] | | Adjusted model[c] | |
|---|---|---|---|---|
| | OR[d] (95% CI) | *P* for trend | AOR[e] (95% CI) | *P* for trend |
| Age (years) | | .26 | | .23 |
| 18 - 34 | 1.26 (0.93 - 1.70) | | 1.20 (0.88 - 1.63) | |
| 34 - 49 | 1.00 (Reference) | | 1.00 (Reference) | |
| 50 - 64 | 1.40 (1.07 - 1.83) | | 1.43 (1.08 - 1.90) | |
| ≥65 | 1.25 (0.64 - 2.47) | | 1.35 (0.67 - 2.72) | |
| Weight status[f] | | —[g] | | |
| Underweight (BMI<18.5 kg/m$^2$) | 0.73 (0.51 - 1.04) | | 0.72 (0.50 - 1.04) | |
| Normal (BMI 18.5 - 24.9 kg/m$^2$) | 1.00 (Reference) | | 1.00 (Reference) | |
| Overweight or obesity (BMI≥25 kg/m$^2$) | 0.97 (0.69 - 1.37) | | 0.95 (0.66 - 1.37) | |
| Missing | 0.89 (0.62 - 1.28) | | 0.94 (0.64 - 1.36) | |
| Sex | | —[g] | | —[g] |
| Male | 1.00 (Reference) | | 1.00 (Reference) | |
| Female | 0.83 (0.64 - 1.07) | | 0.84 (0.63 - 1.11) | |
| Educational level | | .24 | | .26 |
| Junior or senior high school | 1.00 (Reference) | | 1.00 (Reference) | |
| 2-year college or technical school | 1.05 (0.79 - 1.39) | | 1.12 (0.83 - 1.50) | |
| University or higher | 0.79 (0.59 - 1.05) | | 0.78 (0.58 - 1.05) | |
| Others or refusal to answer | 1.18 (0.66 - 2.12) | | 1.25 (0.68 - 2.29) | |
| Employment status | | .02 | | .01 |
| Unemployed | 1.00 (Reference) | | 1.00 (Reference) | |
| Student | 2.12 (0.86 - 5.20) | | 2.12 (0.83 - 5.40) | |
| Part-time job | 1.46 (1.05 - 2.04) | | 1.56 (1.12 - 2.19) | |
| Full-time job | 1.40 (1.04 - 1.90) | | 1.47 (1.07 - 2.02) | |
| Medical history | | —[g] | | —[g] |
| Yes | 0.87 (0.67 - 1.12) | | 0.83 (0.64 - 1.09) | |
| No | 1.00 (Reference) | | 1.00 (Reference) | |
| Refusal to answer | 0.97 (0.56 - 1.67) | | 1.00 (0.57 - 1.74) | |
| Smoking status | | .006 | | .12 |
| Never | 1.00 (Reference) | | 1.00 (Reference) | |
| Past | 1.17 (0.87 - 1.58) | | 0.99 (0.72 - 1.36) | |
| Current (<20 cigarettes a day) | 1.39 (0.98 - 1.98) | | 1.30 (0.90 - 1.90) | |
| Current (≥20 cigarettes a day) | 1.58 (0.83 - 3.00) | | 1.28 (0.65 - 2.50) | |
| Drinking habit | | .07 | | .27 |
| Not at all | 1.00 (Reference) | | 1.00 (Reference) | |



XSL•FO
RenderX

| Variables | Crude model[b] | | Adjusted model[c] | |
|---|---|---|---|---|
| | OR[d] (95% CI) | *P* for trend | AOR[e] (95% CI) | *P* for trend |
| Hardly | 0.68 (0.49 - 0.95) | | 0.67 (0.48 - 0.94) | |
| Sometimes | 0.97 (0.73 - 1.29) | | 0.98 (0.73 - 1.32) | |
| Every day | 1.37 (0.96 - 1.94) | | 1.20 (0.83 - 1.75) | |
| Form of dietary supplement | | —[g] | | —[g] |
| Liquid type | 1.00 (Reference) | | 1.00 (Reference) | |
| Tablet type | 6.12 (3.39 - 11.03) | | 6.67 (3.69 - 12.07) | |
| Featured nutrients of the dietary supplement[h] (tablet type only) | | —[g] | | —[g] |
| Multivitamins/multimineral | 1.00 (Reference) | | 1.00 (Reference) | |
| Folic acid and iron | 0.87 (0.52 - 1.47) | | 0.99 (0.58 - 1.69) | |
| Single fat-soluble vitamins | 1.19 (0.77 - 1.84) | | 1.18 (0.76 - 1.84) | |
| Single water-soluble vitamins | 1.54 (1.13 - 2.10) | | 1.52 (1.11 - 2.09) | |
| Single minerals | 1.00 (0.72 - 1.40) | | 0.95 (0.68 - 1.33) | |
| Duration of the dietary supplement use | | <.001 | | <.001 |
| <3 months | 1.00 (Reference) | | 1.00 (Reference) | |
| 3 months to <6 months | 1.16 (0.76 - 1.77) | | 1.13 (0.74 - 1.72) | |
| 6 months to <1 year | 1.52 (1.04 - 2.22) | | 1.54 (1.06 - 2.23) | |
| ≥1 year | 1.94 (1.42 - 2.65) | | 1.67 (1.22 - 2.30) | |
| Checking the manufacturer-recommended dose before use: “Before you started using the product, did you check the recommended dose of the product?” | | —[g] | | —[g] |
| Yes | 0.53 (0.36 - 0.76) | | 1.20 (0.78 - 1.85) | |
| No | 1.00 (Reference) | | 1.00 (Reference) | |
| Self-perception of their dietary supplement consumption: “Compared to the recommended dose, your consumption is...” | | —[g] | | —[g] |
| More than the recommended dose | 10.77 (6.55 - 17.73) | | 11.46 (6.83 - 19.21) | |
| Same as the recommended dose | 1.00 (Reference) | | 1.00 (Reference) | |
| Less than the recommended dose | 0.89 (0.60 - 1.31) | | 0.89 (0.60 - 1.31) | |
| Unsure | 0.92 (0.67 - 1.26) | | 1.36 (0.97 - 1.90) | |
| No recommended dose provided | 0.21 (0.03 - 1.57) | | 0.37 (0.05 - 2.83) | |
| Food literacy score[i] | | —[g] | | —[g] |
| Less than median score | 1.00 (Reference) | | 1.00 (Reference) | |



XSL•FO
RenderX

| Variables | Crude model[b] | | Adjusted model[c] | |
|---|---|---|---|---|
| | OR[d] (95% CI) | *P* for trend | AOR[e] (95% CI) | *P* for trend |
| Same as median score or more | 1.09 (0.87 - 1.37) | | 1.07 (1.07 - 0.84) | |

[a]Participants were panelists who bought one of the 25 selected dietary supplements in the previous 3 months and used it within the past month or regularly. Information was collected from one of the 25 dietary supplements used by each participant.

[b]Derived from univariate logistic regression models.

[c]Derived from multivariable logistic regression models mutually adjusted for sex, age, weight status, educational level, employment status, medical history, smoking status, drinking habit, and form of dietary supplement; excluded for main variables.

[d]OR: odds ratio.

[e]AOR: adjusted odds ratio.

[f]"Multivitamins/multimineral" contained 3 or more vitamins or minerals. "Folic acid and iron" contained folic acid and iron as the primary nutrients. "Single fat-soluble vitamins" contained <3 fat-soluble vitamins as the primary nutrients. "Single water-soluble vitamins" contained <3 water-soluble vitamins as the primary nutrients. "Single mineral" contained <3 minerals as the primary nutrients.

[g]The *P* for trend was not calculated, since a linear relationship between categories could not be assumedor because only one exposure category (in addition to the reference group) was included..

[h]Users of the liquid-form dietary supplement were excluded from the analysis. The form of the dietary supplement was not adjusted.

[i]The median score among participants was 3.17, with possible scores ranging from 1 to 5.

Table 3 describes the intake of 15 nutrients with UL as defined in the DRIs for Japanese, along with the number of participants taking DSs containing these nutrients. Among 2002 participants, 1705 adults took DS containing ≥1 nutrient with UL. Among 371 adults in the exceeding–manufacturer-recommended-dose group, 297 adults took DS containing ≥1 nutrient with UL. No participant in the within–manufacturer-recommended-dose group exceeded the UL for their DS intake. In the exceeding–manufacturer-recommended-dose group, 61.9% (184/297) exceeded the UL for at least one nutrient. For vitamin A, niacin, folic acid, and zinc, the proportion of users who exceed UL was more than 40% in the exceeding–manufacturer-recommended-dose group using DS containing the nutrient with UL, namely, 50.7% (37/73) for vitamin A, 57% (73/128) for niacin, 48.1% (76/158) for folic acid, 59% (26/44) for magnesium, 52.9% (46/87) for zinc.



XSL•FO
RenderX

**Table .** Nutrient intake with defined UL[a] among dietary supplement users (n=2002)[b].

| Nutrients | UL[c] | All (n=2002) | | Within–manufacturer-recommended dose[d] (n=1631) | | | Exceeding–manufacturer-recommended dose[e] (n=371) | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Participants[f], n | Exceed UL, n (%) | Participant[f]s, n | Nutrient intake, median (IQR) | Exceed UL, n (%) | Participants[f], n | Nutrient intake, median (IQR) | Exceed UL, n (%) |
| Vitamin A (µg) | 2700 | 400 | 37 (9.3) | 327 | 770 (450-770) | 0 (0) | 73 | 3080 (1800 - 5400) | 37 (50.7) |
| Vitamin D (µg) | 100 | 648 | 15 (2.3) | 527 | 6.6 (5.0 - 10.0) | 0 (0) | 121 | 33 (20.0 - 60.0) | 15 (12.4) |
| Vitamin E (mg) | Men: 800<br>Women aged 18-29 years: 650Women aged 30-74 years: 700 | 466 | 3 (0.6) | 381 | 6.3 (6.3 - 10.0) | 0 (0) | 85 | 37.8 (18.9 - 54.0) | 3 (3.5) |
| Niacin[g] (mg) | Men aged 18 - 29 years and 65-74 years: 80<br>Men aged 30 - 64 years: 85<br>Women: 65 | 874 | 73 (8.4) | 746 | 13 (2.1 - 15.0) | 0 (0) | 128 | 80 (57.0 - 104.0) | 73 (57.0) |
| Vitamin $B_6$ (mg) | Men aged 18 - 29 years and 65-74 years: 55<br>Men aged 30 - 64 years: 60<br>Women: 45 | 943 | 39 (4.1) | 805 | 1.3 (0.7 - 3.2) | 0 (0) | 138 | 16 (7.8 - 60.0) | 39 (28.3) |
| Folic acid (µg) | Aged 18 - 29 years, ≥65 years: 900<br>Aged 30 - 64 years: 1000 | 825 | 76 (9.2) | 667 | 200 (100-240) | 0 (0) | 158 | 980 (400 - 1440) | 76 (48.1) |
| Calcium (mg) | 2500 | 270 | 9 (3.3) | 216 | 200 (90-360) | 0 (0) | 54 | 735 (450 - 1500) | 9 (16.7) |
| Magnesium (mg) | 350[h] | 201 | 26 (12.9) | 157 | 125 (100-250) | 0 (0) | 44 | 493 (250-824) | 26 (59.0) |
| Zinc (mg) | Men aged 18 - 29 years: 40<br>Men aged 30 - 74 years: 45<br>Women: 35 | 435 | 46 (10.6) | 348 | 15 (6.0 - 15.0) | 0 (0) | 87 | 45 (18.0 - 90.0) | 46 (52.9) |
| Copper (mg) | 7 | 125 | 2 (1.6) | 94 | 0.6 (0.2 - 0.6) | 0 (0) | 31 | 1.2 (0.8 - 3.6) | 2 (6.5) |
| Manganese (mg) | 11 | 90 | 1 (1.1) | 68 | 1.5 (0.9 - 1.5) | 0 (0) | 22 | 3.3 (2.2 - 6.5) | 1 (4.5) |



XSL•FO
RenderX

| Nutrients | UL[c] | All (n=2002) | | Within–manufacturer-recommended dose[d] (n=1631) | | | Exceeding–manufacturer-recommended dose[e] (n=371) | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Participants[f], n | Exceed UL, n (%) | Participant[f]s, n | Nutrient intake, median (IQR) | Exceed UL, n (%) | Participants[f], n | Nutrient intake, median (IQR) | Exceed UL, n (%) |
| Iodine (µg) | 3000 | 90 | 0 (0.0) | 68 | 45.0 (25.6 - 50.8) | 0 (0) | 22 | 102 (68-162) | 0 (0.0) |
| Selenium (µg) | Men aged 18 - 29 years: 400<br>Men aged 30 - 74 years: 450<br>Women: 350 | 404 | 5 (1.2) | 321 | 50 (30.2 - 50.0) | 0 (0) | 83 | 250 (100-300) | 5 (6.0) |
| Chromium (µg) | 500 | 404 | 4 (1.0) | 321 | 60 (21.4 - 60.0) | 0 (0) | 83 | 180 (85-360) | 4 (4.8) |
| Molybdenum (µg) | Men: 600<br>Women: 500 | 90 | 0 (0.0) | 68 | 9.0 (5.2 - 10.5) | 0 (0) | 22 | 21 (14.0 - 32.4) | 0 (0.0) |
| ≥1 nutrients above | | 1705 | 184 (10.8) | 1416 | — | 0 (0) | 297 | — | 184 (61.9) |

[a]UL: tolerable upper intake level.

[b]Participants were panelists who had bought one of the 25 selected dietary supplements in the last 3 months and used it within the past month or regularly. Information was collected from one of the 25 dietary supplements used by each participant.

[c]UL is defined by the Dietary Reference Intakes for Japanese, 2025.

[d]Those whose daily intake of the dietary supplement was equal to or below the manufacturer-recommended amount.

[e]Those whose intake exceeded the manufacturer-recommended amount.

[f]Number of participants taking nutrient-containing dietary supplements in the leftmost column.

[g]Amount of nicotinamide.

[h]ULs were defined for intakes derived from dietary supplements only.

For sensitivity analysis, 10.4% (190/1821) of patients were categorized into the exceeding–manufacturer-recommended-dose group. The results for the factors associated with DS consumption exceeding the manufacturer-recommended doses were almost unchanged from the original analysis of the 2002 participants (Tables S4 and S5 in Multimedia Appendix 1). The number of participants who consumed DS containing ≥1 nutrients with ULs was 134 adults. Of these, the number of participants who exceeded UL for at least one nutrient with ULs was 44 (32.8%). These numbers were much smaller than those included in the main analysis (Table S6 in Multimedia Appendix 1).

## Discussion

### Principal Findings

To the best of our knowledge, this is the first study to examine the factors associated with DS consumption exceeding the manufacturer-recommended doses among Japanese adults. Among 2002 DS users, 371 (18.5%) consumed DS that exceeded the manufacturer-recommended doses indicated on the package. Consumption of DS above the recommended dose was associated with middle age, part-time or full-time employment, use of tablet-form DS, especially single water-soluble vitamin tablets, use of DS for 6 months or more, and conscious consumption above the recommended dose. Among 1705 individuals consuming DS containing nutrients with ULs, 17.4% (n=297) exceeded the recommended dose. Of these, 61.9% (184/297) of individuals surpassed the UL by at least one nutrient. The results will contribute to a more accurate understanding of DS overuse behaviors, their associated factors, and possible health outcomes.

In this study, the proportion of participants who exceeded the manufacturer-recommended dose was 371 (18.5%) out of 2002, which is substantially higher than that previously reported (2.4%) [20]. When the proportion was calculated based on the self-perception of DS consumption, 4.1% ([25+57]/2002) exceeded the manufacturer-recommended dose, which is close to that previously reported. Among the 371 participants, only 15.4% (57/371) were aware that their consumption exceeded the recommended dose. These findings suggest that self-perception of adherence to the recommended dose is not a reliable indicator of actual dose consumption. To accurately assess compliance with the recommended doses, identifying the specific product used and comparing the reported consumption with the corresponding manufacturer-recommended dose are essential.

Previous research on DS overuse has focused on the use of multiple DS products [28,29], and to our knowledge, no studies have specifically examined DS consumption exceeding the manufacturer-recommended dose. Therefore, although the reasons for the observed associations with the abovementioned factors remain speculative, several possible explanations can



XSL•FO
RenderX

be considered. Among older adults, a higher prevalence of multiple DS use has been reported in the US population [28,29]. As health deterioration associated with aging may increase interest in diet [30], older individuals may consciously or unconsciously consume higher amounts of DS. Regarding employment status, individuals with part-time or full-time jobs may rely on DS as a convenient means of supplementing their nutrition due to busy schedules, which could lead to excessive DS intake. A previous study has reported that individuals working long hours often perceive barriers to consuming healthy meals [31]. Since only 31.4% (627/2002) of participants reported a medical history in this study, and 70% of DS users among patients did not disclose DS use to their physicians in a previous study [32], it was unlikely that participants had exceeded the manufacturer-recommended dose of DS based on specialists' recommendations.

Users of the DS tablet form, especially those using a single water-soluble vitamin tablet, were more likely to exceed the manufacturer-recommended dose. In addition, use of DS for 6 months or more, and conscious consumption above the recommended dose were also associated. With respect to the form of DS, liquid DS may be used more episodically and may be less likely to result in use above the recommended dose, as a previous study suggested that energy drinks are used episodically rather than regularly among college students [33]. According to a US study, DS of water-soluble vitamins were used to expect that it would boost the immune system, prevent colds, and improve overall health [34]. Some participants may have intentionally consumed more than the manufacturer-recommended dose regularly, believing that it would yield greater health benefits, even when they were aware of the manufacturer-recommended dosages. Moreover, a considerable proportion of individuals may have either misinterpreted or failed to retain the correct information regarding the recommended intake even after reviewing it. For such individuals, even if dosage guidelines are clearly indicated on labels, the information may not be effectively used. Notably, no significant differences in food literacy scores were found between those who adhered to appropriate dosage levels and those who consumed higher dosage levels. This finding implies that general food literacy scales may not adequately capture the literacy specific to DS use. Alternatively, factors beyond literacy, such as health beliefs, perceived efficacy, and trust in DS, may contribute to excessive intake. To prevent overconsumption of DSs, it may be necessary to improve the clarity and accessibility of dosage-related information and increase public awareness of the potential harm associated with excessive intake.

In this study, among 1705 participants who consumed DS containing nutrients with UL, 10.7% (184/1705) of participants exceeded the UL for at least one nutrient. They were asked about their consumption of only one DS product. Although some participants may have obtained nutrients from other DS sources, this was not captured in the survey. Previous studies in the United States reported that the concurrent use of multiple DS was common [28,29]. For example, middle-aged and older DS users use an average of more than 3.2 different DS, with 41.9% using 4 or more products [28]. Another US study on adults found that all individuals whose usual daily folic acid intake exceeded the UL consumed folic acid from multiple sources [16]. Consequently, among our study participants, the proportion of individuals exceeding the UL for specific nutrients may have been higher if nutrient intake from multiple concurrently used DS had been considered.

Previous studies have reported the proportion of individuals exceeding the UL. In the United States, approximately 20% - 40% of adults use vitamin DSs, with reported rates of exceeding UL at 7% for folic acid, 3% for vitamin A, 3.5% for vitamin $B_6$, and 1.6% for vitamin C [13]. Approximately 7% - 12% of DS users exceeded the UL for iron or zinc, and less than 3% - 7% exceeded the UL for calcium or magnesium [11]. In the Netherlands, although the nutrient intake from both DSs and fortified foods exceeded the UL for some participants, the proportions were <3% for calcium, iron, vitamins $B_6$, vitamin A, and folate [14]. In Canada, when intake from DS was included, ≥10% of individuals in certain age and sex groups had intakes of vitamins A and C, niacin, folic acid, iron, zinc, and magnesium that exceeded the UL [15]. However, few studies have provided detailed descriptions of how individuals use DS, such as whether they use multiple DS concurrently or whether they consume doses exceeding the manufacturer-recommended dose. Therefore, future studies should not only report the proportion of individuals exceeding the UL but also examine the behavioral patterns of DS use to better understand the factors contributing to excessive nutrient intake.

## Strengths and Limitations

The strength of this study lies in the use of purchase history data to recruit participants. We targeted panelists who purchased and used specified DS products. This approach helped minimize misreporting related to product identification, such as when respondents were unable to recall or report product names accurately. In addition, presenting the specific unit for the product in the questionnaire might reduce the misreporting of the dosage, although some misreporting appeared. Consequently, this enabled a reliable comparison between the participants' reported DS consumption and the recommended dosage for each product. However, this study has some limitations. First, the participants were limited to panelists registered as online survey monitors, which may compromise the generalizability of the findings owing to potential sampling bias. Our participants had higher education levels than the nationally representative sample aged 20 - 74 years (43.2% for junior or senior high school, 15.6% for 2-year college or technical school, 24.2% for university or higher; education level reported in the Population Census of Japan 2020 [35]). Second, the extent to which the DS included in this study represents the full range of products consumed by Japanese DS users remains unclear and may be due to the lack of comprehensive and nationwide statistical data on the consumption of DS products. However, as the items were selected based on their high purchase frequency in transaction records, they were considered to ensure a certain level of representativeness. Third, data on daily DS consumption and frequency of consumption were self-reported, which introduces the possibility of reporting errors. Although the survey was



XSL•FO
RenderX

designed to facilitate accurate responses by clearly specifying the units of each product on the response screen, some participants likely misreported their consumption quantities. We excluded those who consumed extremely high doses of DS (≥360 tablets per day) from the analysis because they were likely to have overreported it; however, some participants in the analyzed group may have also overreported it. However, when we excluded participants who consumed 5 times the manufacturer-recommended doses, the results were similar (Table S1 in Multimedia Appendix 1). Fourth, this study did not assess participants' concurrent use of multiple DS. To ensure the feasibility of the survey and obtain accurate responses regarding the selected products, participants with purchase histories of multiple DS products were randomly assigned to answer questions about only one of those products. Consequently, information on the use of multiple DS and their contributions to overall nutrient intake and the adequacy of nutrient intake remains unknown. If the concurrent use of multiple DS is also considered, the proportion of participants exceeding the UL may increase further, indicating the need for greater public awareness and caution regarding DS use. Further studies are required to better understand the patterns of multiple DS use and their implications for total nutrient consumption.

## Conclusions

In conclusion, middle-aged, part-time or full-time jobs, using tablet-form DS, and using DS for 6 months or more were associated with DS consumption exceeding the manufacturer-recommended dose, which may lead to excessive nutrient intake. Some individuals consume excessive doses of DSs, highlighting the need to raise awareness about the potential health risks associated with overconsumption.

## Acknowledgments

The authors would like to express their sincere gratitude to the research staff of the University of Tokyo for their assistance in extracting nutrient content data from the manufacturer's website. The authors declare the use of generative AI (GAI) in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GAI tools under full human supervision: proofreading and editing. The GAI tool used was ChatGPT 5.1. Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes.

## Funding

This study was financially supported by a Food Safety Standards Sciences Research grants (23KA1009) from the Consumer Affairs Agency, Japan. The Consumer Affairs Agency had no role in the design, analysis, or writing of this article.

## Data Availability

The datasets generated and analyzed during this study are not publicly available but may be made available by the corresponding author on reasonable request.

## Authors' Contributions

Conceptualization: MS, KA
Data curation: MS, KA
Formal analysis: MS, KA
Funding acquisition: KM
Investigation: KA
Methodology: MS
Project administration: KA
Software: MS
Supervision: KA
Visualization: MS, KA
Writing – original draft: MS
Writing – review & editing: KA, NS, KM

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional tables describing analyses of products and participants, along with sensitivity analyses.
[DOCX File, 129 KB - i-jmr_v15i1e82623_app1.docx ]

## References



XSL•FO
RenderX

1. Djaoudene O, Romano A, Bradai YD, et al. A Global overview of dietary supplements: regulation, market trends, usage during the COVID-19 pandemic, and health effects. Nutrients 2023 Jul 26;15(15):3320. [doi: 10.3390/nu15153320] [Medline: 37571258]
2. Kantor ED, Rehm CD, Du M, White E, Giovannucci EL. Trends in dietary supplement use among US adults from 1999-2012. JAMA 2016 Oct 11;316(14):1464-1474. [doi: 10.1001/jama.2016.14403] [Medline: 27727382]
3. Zhao L, Zhang Y, Liu J, et al. Trends in dietary supplement use among U.S. adults between 2011 and 2023. Eur J Nutr 2025 Oct 23;64(8):30. [doi: 10.1007/s00394-025-03825-4] [Medline: 41128907]
4. Dietary supplements: information on consumption of dietary supplements (including fibre and protein) in Australia. Australian Bureau of Statistics. 2025 Sep 5. URL: https://www.abs.gov.au/statistics/health/food-and-nutrition/dietary-supplements/latest-release [accessed 2026-02-25]
5. Baker B, Probert B, Pomeroy D, Carins J, Tooley K. Prevalence and predictors of dietary and nutritional supplement use in the Australian Army: a cross-sectional survey. Nutrients 2019 Jun 27;11(7):1462. [doi: 10.3390/nu11071462] [Medline: 31252600]
6. Flynn A, Hirvonen T, Mensink GBM, et al. Intake of selected nutrients from foods, from fortification and from supplements in various European countries. Food Nutr Res 2009;53. [doi: 10.3402/fnr.v53i0.2038] [Medline: 20011225]
7. Sugimoto M, Asakura K, Shinozaki N, Murakami K, Masayasu S, Sasaki S. Contribution of fortified foods and dietary supplements to total nutrient intakes and their adequacy in Japanese adults. BMC Nutr 2024 Sep 27;10(1):125. [doi: 10.1186/s40795-024-00935-w] [Medline: 39334410]
8. Dietary Reference Intakes for Japanese (2025 Edition). Ministry of Health, Labour and Welfare. 2025. URL: https://www.mhlw.go.jp/stf/newpage_44138.html [accessed 2026-02-25]
9. Bailey RL, Fulgoni VL III, Keast DR, Dwyer JT. Dietary supplement use is associated with higher intakes of minerals from food sources. Am J Clin Nutr 2011 Nov;94(5):1376-1381. [doi: 10.3945/ajcn.111.020289] [Medline: 21955646]
10. Fassier P, Egnell M, Pouchieu C, et al. Quantitative assessment of dietary supplement intake in 77,000 French adults: impact on nutritional intake inadequacy and excessive intake. Eur J Nutr 2019 Oct;58(7):2679-2692. [doi: 10.1007/s00394-018-1815-x] [Medline: 30293178]
11. Blumberg JB, Frei BB, Fulgoni VL, Weaver CM, Zeisel SH. Impact of frequency of multi-vitamin/multi-mineral supplement intake on nutritional adequacy and nutrient deficiencies in U.S. adults. Nutrients 2017 Aug 9;9(8):849. [doi: 10.3390/nu9080849] [Medline: 28792457]
12. Kiely M, Flynn A, Harrington KE, et al. The efficacy and safety of nutritional supplement use in a representative sample of adults in the North/South Ireland Food Consumption Survey. Public Health Nutr 2001 Oct;4(5A):1089-1097. [doi: 10.1079/phn2001190] [Medline: 11820922]
13. Bailey RL, Fulgoni VL III, Keast DR, Dwyer JT. Examination of vitamin intakes among US adults by dietary supplement use. J Acad Nutr Diet 2012 May;112(5):657-663. [doi: 10.1016/j.jand.2012.01.026] [Medline: 22709770]
14. Bird JK, Bruins MJ, Turini ME. Micronutrient intakes in the Dutch diet: foods, fortified foods and supplements in a cross sectional study. Eur J Nutr 2023 Dec;62(8):3161-3179. [doi: 10.1007/s00394-023-03219-4] [Medline: 37542641]
15. Shakur YA, Tarasuk V, Corey P, O'Connor DL. A comparison of micronutrient inadequacy and risk of high micronutrient intakes among vitamin and mineral supplement users and nonusers in Canada. J Nutr 2012 Mar;142(3):534-540. [doi: 10.3945/jn.111.149450] [Medline: 22298574]
16. Orozco AM, Yeung LF, Guo J, Carriquiry A, Berry RJ. Characteristics of U.S. adults with usual daily folic acid intake above the tolerable upper intake level: National Health and Nutrition Examination Survey, 2003-2010. Nutrients 2016 Apr 1;8(4):195. [doi: 10.3390/nu8040195] [Medline: 27043623]
17. Burt LA, Billington EO, Rose MS, Raymond DA, Hanley DA, Boyd SK. Effect of high-dose vitamin D supplementation on volumetric bone density and bone strength: a randomized clinical trial. JAMA 2019 Aug 27;322(8):736-745. [doi: 10.1001/jama.2019.11889] [Medline: 31454046]
18. Dietary supplement labeling guide: Chapter I. General dietary supplement labeling. U.S. Food and Drug Administration. 2005 Apr. URL: https://www.fda.gov/food/dietary-supplements-guidance-documents-regulatory-information/dietary-supplement-labeling-guide-chapter-i-general-dietary-supplement-labeling [accessed 2026-02-25]
19. Directive 2002/46/EC of the European Parliament and of the Council of 10 June 2002 on the approximation of the laws of the Member States relating to food supplements. Off J Eur Communities L 2002;183:51-57 [FREE Full text]
20. Survey on the actual use of "health foods" by consumers (questionnaire survey). Consumer Affairs Agency. 2012 May. URL: https://www.cao.go.jp/consumer/iinkaikouhyou/2012/houkoku/201205_report.html [accessed 2026-02-25]
21. Bailey RL, Dodd KW, Gahche JJ, et al. Best practices for dietary supplement assessment and estimation of total usual nutrient intakes in population-level research and monitoring. J Nutr 2019 Feb 1;149(2):181-197. [doi: 10.1093/jn/nxy264] [Medline: 30753685]
22. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet e-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
23. Guidelines for labeling intake amount and method of so-called health foods. : Ministry of Health, Labour and Welfare; 2005 Feb 28 Notification no. 0228001 URL: https://www.mhlw.go.jp/topics/bukyoku/iyaku/syoku-anzen/topics/050131/dl/tuuchi08.pdf [accessed 2026-02-25]




XSL•FO
RenderX

24. Guidelines for notification of foods with function claims [Article in Japanese]. : Consumer Affairs Agency; 2023 Sep 29 Notification no. 543 URL: https://www.caa.go.jp/policies/policy/food_labeling/foods_with_function_claims/assets/foods_with_function_claims_230929_0002.pdf [accessed 2026-02-25]
25. Approval standards for newly designated quasi-drugs [Article in Japanese]. : Ministry of Health, Labour and Welfare; 1999 Mar 12 URL: https://www.mhlw.go.jp/file/06-Seisakujouhou-11120000-Iyakushokuhinkyoku/sinsitei.pdf [accessed 2026-03-13]
26. Poelman MP, Dijkstra SC, Sponselee H, et al. Towards the measurement of food literacy with respect to healthy eating: the development and validation of the self perceived food literacy scale among an adult sample in the Netherlands. Int J Behav Nutr Phys Act 2018 Jun 18;15(1):54. [doi: 10.1186/s12966-018-0687-z] [Medline: 29914503]
27. Murakami K, Shinozaki N, Okuhara T, McCaffrey TA, Livingstone MBE. Self-perceived food literacy in relation to the quality of overall diet and main meals: a cross-sectional study in Japanese adults. Appetite 2024 May 1;196:107281. [doi: 10.1016/j.appet.2024.107281] [Medline: 38373536]
28. Tan ECK, Eshetie TC, Gray SL, Marcum ZA. Dietary supplement use in middle-aged and older adults. J Nutr Health Aging 2022;26(2):133-138. [doi: 10.1007/s12603-022-1732-9] [Medline: 35166304]
29. Gahche JJ, Bailey RL, Potischman N, Dwyer JT. Dietary supplement use was very high among older adults in the United States in 2011-2014. J Nutr 2017 Oct;147(10):1968-1976. [doi: 10.3945/jn.117.255984] [Medline: 28855421]
30. Becker CM, Arnold W. Health promoting behaviors of older Americans versus young and middle aged adults. Educ Gerontol 2004 Nov;30(10):835-844. [doi: 10.1080/03601270490507277]
31. Escoto KH, Laska MN, Larson N, Neumark-Sztainer D, Hannan PJ. Work hours and perceived time barriers to healthful eating among young adults. Am J Health Behav 2012 Nov;36(6):786-796. [doi: 10.5993/AJHB.36.6.6] [Medline: 23026037]
32. Chiba T, Sato Y, Nakanishi T, Yokotani K, Suzuki S, Umegaki K. Inappropriate usage of dietary supplements in patients by miscommunication with physicians in Japan. Nutrients 2014 Nov 26;6(12):5392-5404. [doi: 10.3390/nu6125392] [Medline: 25431879]
33. Malinauskas BM, Aeby VG, Overton RF, Carpenter-Aeby T, Barber-Heidal K. A survey of energy drink consumption patterns among college students. Nutr J 2007 Oct 31;6:35. [doi: 10.1186/1475-2891-6-35] [Medline: 17974021]
34. Bailey RL, Gahche JJ, Miller PE, Thomas PR, Dwyer JT. Why US adults use dietary supplements. JAMA Intern Med 2013 Mar 11;173(5):355-361. [doi: 10.1001/jamainternmed.2013.2299] [Medline: 23381623]
35. 2020 Population Census: Basic tabulation of employment status, etc. (Main contents: Labor force status, industry and occupation of employed persons, education, etc.). Statistics Bureau of Japan. 2022 May. URL: https://www.e-stat.go.jp/stat-search/files?page=1&stat_infid=000032201217 [accessed 2026-02-25]


## Abbreviations

**AOR:** adjusted odds ratio
**CHERRIES:** Checklist for Reporting Results of Internet e-Surveys
**DRI:** Dietary Reference Intakes for Japanese
**DS:** dietary supplement
**UL:** tolerable upper intake level


*Edited by M Balcarras; submitted 19.Aug.2025; peer-reviewed by A Woźniak, G Passeri, MA Rifat, R Wierzejska; revised version received 09.Feb.2026; accepted 11.Feb.2026; published 19.Mar.2026.*

*Please cite as:*
*Sugimoto M, Asakura K, Shinozaki N, Murakami K*
*Prevalence and Associated Factors of Excessive Dietary Supplement Use Among Japanese Adults: Cross-Sectional Study*
*Interact J Med Res 2026;15:e82623*
*URL: https://www.i-jmr.org/2026/1/e82623*
*doi:10.2196/82623*


© Minami Sugimoto, Keiko Asakura, Nana Shinozaki, Kentaro Murakami. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 19.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Determining the Relationship Between People's Explicit and Implicit Preferences for Gender-Inclusive Sexual and Reproductive Health Content: Randomized Controlled Trial

Elizabeth R Boskey[1,2,3], MSSW, MPH, PhD; Jessica D Kant[1], MSW, MPH; Ariel K Berman[4,5], PhD; Frances W Grimstad[1,6], MS, MD

[1]Division of Gynecology, Department of Surgery, Boston Children's Hospital, 300 Longwood Ave, Boston, MA, United States

[2]Department of Surgery, Medical School, Harvard University, Boston, MA, United States

[3]Department of Epidemiology, T.H. Chan School of Public Health, Harvard University, Boston, MA, United States

[4]Department of Psychiatry and Behavioral Sciences, Boston Children's Hospital, Boston, MA, United States

[5]Department of Psychiatry, Medical School, Harvard University, Boston, MA, United States

[6]Department of Obstetrics, Gynecology, and Reproductive Biology, Medical School, Harvard University, Boston, MA, United States

**Corresponding Author:**
Elizabeth R Boskey, MSSW, MPH, PhD
Division of Gynecology, Department of Surgery, Boston Children's Hospital, 300 Longwood Ave, Boston, MA, United States

## Abstract

**Background:** Inclusive health education content has been shown to increase acceptability and accessibility for lesbian, gay, bisexual, transgender, queer, intersex, and asexual, as well as other sexual and gender minority (LGBTQ+) individuals. However, there has been some backlash among general audiences, with claims that such inclusive content is "woke" or otherwise problematic.

**Objective:** The goal of this study was to test whether individuals across the political spectrum notice when sexual and reproductive health content is written with inclusive language in order to demonstrate the acceptability of inclusive content to a broader audience.

**Methods:** This study included 454 adults assigned female at birth from the United States, one-third of whom identified as LGBTQ+, reviewed 2 sets of reproductive health educational handouts designed for adolescents, with 1 gender-inclusive and 1 gender-specific version in each set, randomized in order. Individuals were asked to rate each document and state a preference within each pair (implicit preference). They were then debriefed on the study's purpose and asked if they had an explicit preference for gender-specific or gender-inclusive content.

**Results:** Preferences for explicit content tended toward gender-specific content: always gender-specific (n=184, 40.5%), sometimes gender-specific (n=59, 13%), no preference (n=131, 28.8%), sometimes gender-inclusive (n=39, 8.6%), and always gender-inclusive (n=41, 9%). However, most people (n=273, 59%) did not notice differences between the first pair of documents they viewed or rate them differently (mean difference −0.19, SD 2.17, range −10 to 12). Furthermore, the majority of individuals who had a stated preference for gender-specific health education documents did *not* choose the gender-specific document as their preferred version for either the first (n=45, 24%) or second pair of documents (n=69, 38%). Individuals who preferred content to always be gender-inclusive were significantly more likely to choose the concordant version of their document (n=20, 49% for the first pair; n=24, 58% for the second pair). A total of 58% (n=262) of the participants stated they did not notice the study design until the debrief.

**Conclusions:** Most participants did not notice when sexual and reproductive health educational content had been made gender-inclusive—even when they had an explicit preference for gender-specific content. This suggests that when inclusive language is not directly called to readers' attention, inclusive sexual and reproductive health content is broadly acceptable to individuals across a range of political beliefs. The use of inclusive language may therefore be a means of increasing the accessibility and applicability of educational materials to diverse recipients, including LGBTQ+ individuals.

**Trial Registration:** ClinicalTrials.gov NCT07601581; https://clinicaltrials.gov/study/NCT07601581

***(Interact J Med Res 2026;15:e85868)*** doi:10.2196/85868

**KEYWORDS**

inclusive language; sexual and reproductive health; health education; randomized controlled trial; sex factors; gender identity; sex education



XSL•FO
RenderX

## Introduction

In the early 2000s, as acceptance for lesbian, gay, bisexual, transgender, queer, intersex, and asexual, as well as other sexual and gender minority (LGBTQ+) individuals was growing across the population of the United States, there was increasing awareness of the need to create sexual and reproductive health education content that would be more inclusive of sexuality and gender [1-3]. One way to accomplish this is by enhancing patient educational materials through reducing or removing assumptions about the gender identity of individuals being discussed, and for sexual health materials, also removing assumptions about the genders of discussed sexual partners, that is, making the content inclusive. This approach has also been described as degendering health content. Over time, language options were developed for these materials to be more inclusive of a broad range of gender identities and sexual orientations (eg, pregnant person, birthing parent, lactating parent, sperm-producing partner, person with a uterus, ovaries, vagina, penis, etc). This often involved focusing on anatomy or experience rather than identity, as those aspects of life tended to be more relevant to the health question or behavior of interest [4]. Inclusive language makes sexual and reproductive health content more relevant to and actionable for lesbian, gay, bisexual, and transgender individuals rather than reinforcing notions that these aspects of a healthy life are only accessible and attainable to those who are cisgender and heterosexual [5-7].

The development of inclusive content has frequently been accompanied by opposition, often politically and/or religiously driven [8-10]. This opposition is often based on inaccurate beliefs that all individuals who have the same anatomy (eg, uterus) will have the same gender (eg, woman). This is problematic not just because up to 2% of the population identify as transgender [11], but because other women may have been born without a uterus or had it removed. In recent years, there has been an exponential growth in attacks on gender and sexual diversity across media and government [12-14]. This has peaked under the second Trump administration, with numerous executive orders and policy initiatives designed to erase transgender and other gender diverse individuals from public life [15-17]. There is also sometimes a perception that it is necessary to safeguard women-centric health care with women-centric language [18,19]. This is particularly true in obstetrics and gynecology—even though not all aspects of women's health are nested within obstetrics and gynecology and not all patients who see obstetricians and gynecologists are women

For many individuals whose sexuality, anatomy, and/or gender differ from expected binary norms, including those who may have been born with variations in their sexual and reproductive anatomy (eg, without a uterus) or have undergone surgeries to alter it (eg, a hysterectomy), health education content that assumes gender based on anatomy, or anatomy based on gender, may not come across as accurate, relevant, or useful, and thus fail to meet their needs. As such, it may be helpful to develop sexual and reproductive health content that is anatomy and behavior-specific and not reliant upon assumptions about gender or sexuality [20]. However, that usefulness is dependent on inclusive medical content not being off-putting, offensive, or considered unrelatable, to heterosexual individuals whose anatomy and gender do reflect binary norms.

In response to community calls for greater uptake of inclusive content, some individuals have expressed strong feelings and beliefs about the dangers or offensiveness of such content [21]. However, to date, no studies have examined whether those opposed to inclusive language notice when such language is present but not specifically brought to their attention. This randomized controlled trial was therefore designed to test the hypothesis that most individuals do not notice when sexual and reproductive content has been changed to use inclusive language, with the exception of those for whom inclusive language is more personally relevant. It was also designed to test whether political identity and/or LGBTQ+ identity were associated with explicit preferences for gender-specific versus inclusive content and/or the likelihood of noticing and preferring such content.

## Methods

### Study Population

The Prolific survey platform [22] was used to recruit individuals assigned female at birth who lived in the United States for a study in which they were informed that they would be shown 2 different versions of online health content and asked which they preferred. In order to ensure respondents are genuine, potential participants must provide a photo taken with a state or federal ID when registering for the platform. The site validates IP addresses to prevent duplication, uses machine learning to screen for potential automation, and allows studies to withhold payment for suspicious responses, including those completed substantially faster than the average time for completion.

The Prolific quota function was used to recruit a population that was approximately half LGBTQ+-identified and equally divided among conservative, moderate, and liberal political orientations based on the site's built-in demographic questionnaires. Individuals were also asked to self-identify within the survey, and those identifications were used in the analysis. Because of this, the actual sample demographics were not identical to the planned quota sample, reflecting the fact that individuals may have completed initial screeners for the Prolific platform up to several years prior to any given survey participation and some of their answers may have changed.

### Study Design

Individuals were asked to review 2 versions of 2 sets of patient-educational materials that were originally designed for adolescents assigned female at birth—the first on polycystic ovarian syndrome and the second on birth control. Within each set, 1 document was made inclusive, while the other used gender-specific language. Two study team members reviewed language differences in the documents to ensure consensus on language choice. The documents were otherwise identical, except that one (redundant) figure in the inclusive birth control document was removed due to the use of images of pregnant women in the figure (changes are listed in Table 1). Individuals were randomized using an allocation table generated in Excel,



XSL•FO
RenderX

so that half received the inclusive document first in each set, while the other half received the explicitly gendered material first.

**Table .** Text changes made from gender-specific to gender-inclusive document versions.

| Context for change (if needed) | Gender-specific version | Gender-inclusive version |
|---|---|---|
| PCOS[a] (~1550 words) | | |
| PCOS is a common problem among… | • girls and young women | • people who have ovaries |
| Almost 1 out of 10… | • women have PCOS | • people who have ovaries have PCOS |
| Section Heading: | • Female reproductive anatomy | • Reproductive anatomy of someone with a uterus and ovaries |
| The ovaries make estrogen and progesterone… | • (the female sex hormones) | • [deleted] |
| 3x in document | • women and girls with PCOS | • people with PCOS |
| 10x in document | • women | • people |
| … but if you have PCOS, your ovaries make a little bit more testosterone than they are supposed to | • All women make testosterone | • Everyone makes testosterone |
| • -- | • teenage girls | • adolescents |
| Birth control (~1400 words) | | |
| -- | • female-controlled contraception | • contraception |
| These are similar to the estrogen and progesterone normally made by [] | • women's ovaries | • the ovaries |
| • -- | • female hormones | • estrogen and progestin |
| • the male's sperm | • sperm | |
| some [] can't take or tolerate estrogen | • girls and women | • people |
| 8x in document | • women | • people |
| -- | • condoms | • condoms or other barrier methods |
| Perfect use can be difficult for… | • teen girls and adult women | • teens and adults |
| Very occasionally, your breasts may become tender and/or get larger, but usually your breasts will stay the same. [] | • Breast tenderness | • This tenderness |
| Can any [] take birth control pills? | • woman | • person born with a uterus and ovaries |
| Almost all [] can take birth control pills | • women | • people born with a uterus and ovaries |
| Table describing pregnancy rates associated with birth control pill use | • Has images of pregnant silhouettes as illustrations of the number of pregnancies, alongside text | • Text only |

[a]PCOS: polycystic ovary syndrome, name since changed to PMOS (polyendocrine metabolic ovarian syndrome).

After reading each document, individuals were asked to rate the content on a 5-point Likert scale (−2 to +2) across the following domains: easy to understand, reliable, biased, meant for people like me, and relevant to people like me. These scores were then summed to provide an overall score for each document, with the bias question reverse scored. Individuals were also allowed to comment on each rating. After reviewing each pair of documents, individuals were asked if they had a version preference and to explain their reasoning. A separate comparison score was generated by subtracting individuals' gender-inclusive score from the gender-specific score, so that higher comparison scores were associated with stronger implicit preferences for the gender-specific version.



XSL•FO
RenderX

Finally, after reviewing both sets of informational material, individuals completed a demographic survey that concluded with a question about whether sex education should be required as part of public K-12 education. They were then presented with a debrief that explained the purpose of the study, which included asking whether they had noticed that they were being presented with gender-inclusive material and if they preferred material to be gender-specific or gender-inclusive (explicit preference). Individuals were also given the opportunity to provide their closing thoughts. The questions are provided as a supplementary document. Individuals were not able to return to previous sections of the study or revise their answers to questions. All materials were accessed via Research Electronic Data Capture [23,24].

### Ethical Considerations

This study was approved by the Boston Children's Hospital IRB (IRB-P00047950).

### Analysis

#### *Power Calculation*

The study was powered based on the use of a parallel 2-group design to test whether group A (gender-inclusive first) mean (μ1) was equivalent to group B (gender-specific first) mean (μ2), with difference equivalence bounds of −1 and 1 (H0: $\delta \leq -1$ or $\delta \geq 1$ vs H1: $-1<\delta<1$, $\delta = \mu1-\mu2$), with an overall type I error rate (α) of .05. The common SD was assumed to be the fixed value of 2, with the actual SD for the 2 comparison scales ranging from 2.2 to 2.3. To detect a mean difference ($\delta = \mu1-\mu2 = \mu T-\mu R$) of 0 with sample sizes of 225 for group 1 (treatment) and 225 for group 2 (reference), the power was .9997. With the actual sample breakdown of 176 in the LGBT+ group and 278 in the cisgender-heterosexual group, the power was calculated as .9996 to determine equivalence between those groups. The power was computed using PASS 2023 (version 23.0.7).

#### *Quantitative Analysis*

Quantitative analyses were performed in Stata 18.0 (StataCorp LLC). Demographic variables are presented as numbers and percentages. Differences across categories of explicit, stated preferences for gender-inclusive information were calculated using the chi-square test.

For each document, pair preferences for those with varying explicit preferences were categorized as concordant (ie, states they prefer gender-specific information, chose gender-specific document as preferred) or discordant or do not notice (ie, either stated there was no difference between the documents or chose the type of document they said they did not prefer). Differences across explicit preference categories were calculated using the chi-square test as well as for when individuals stated they noticed that 1 document in each pair was gender-inclusive. Differences in calculated preference scores for each pair of documents were analyzed using ANOVA.

#### *Qualitative Analysis*

Open-ended answers were analyzed in the Dedoose cloud app for managing, analyzing, and presenting qualitative and mixed method research data (Version 9.2.22: Sociocultural Research Consultants, LLC). Excerpts were excluded from analysis if they did not provide additional information (ie, were a repeat of the answer from a numeric field, or contained a one-word answer such as "no," "yes," or "useful"). They were also excluded if they stated a fact unrelated to the question of interest (eg, "This content is not relevant to me because I do not use birth control pills"). Due to the brevity of responses, inductive content analysis was used to identify commonalities between comments, as well as whether this varied by politics and/or stated (explicit) preference for gender-inclusive information [25,26]. Codes were first generated inductively and then grouped by frequency to provide an overview of types of responses. Summative or frequency information is additionally provided to demonstrate between-group differences.

## Results

### Overview

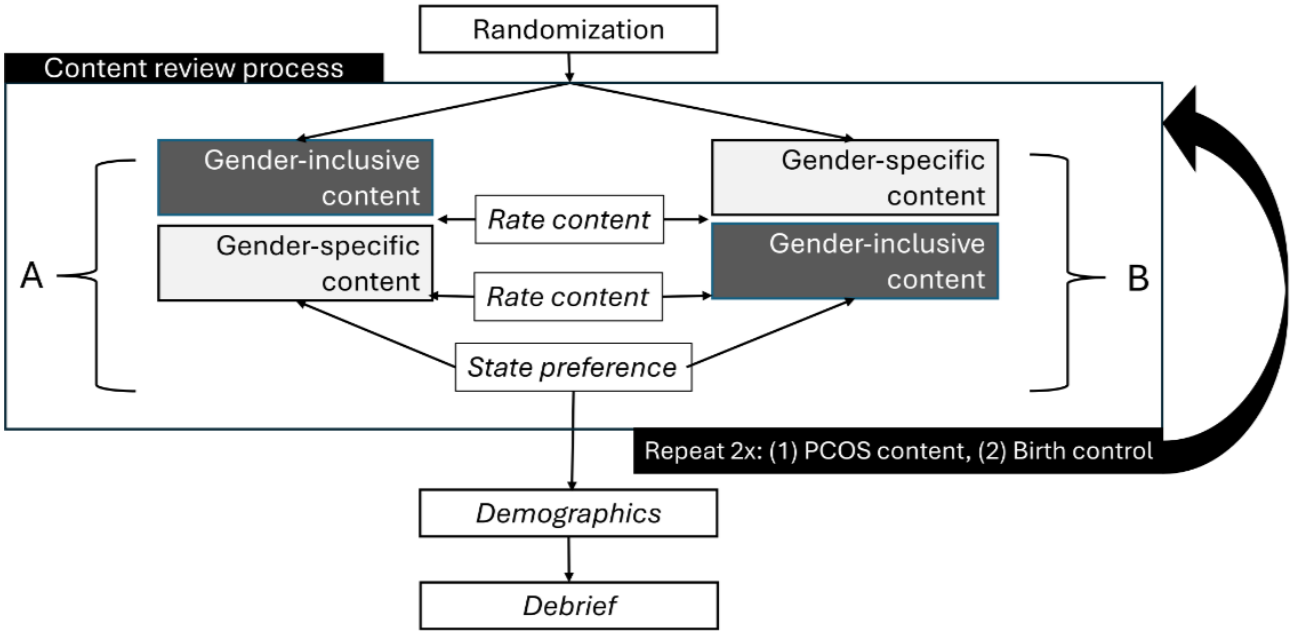
In total, 494 individuals participated in the study, 227 in the version that received the gender-inclusive documents first and 227 in the version that received the gender-specific documents first. There were 40 people who started the study but did not complete it, resulting in a response or completion rate of 92% (n=454). The study flow is outlined in Figures 1 and 2.



XSL•FO
RenderX

**Figure 1.** Study flow design. Each assessment is listed in italics. Simultaneously attempting to make quotas on both LGBT identity (1:1) and political identity (1:1:1) led to slightly unbalanced samples, although with sufficient diversity for analyses. LGBT: lesbian, gay, bisexual, transgender; PCOS: polycystic ovary syndrome.

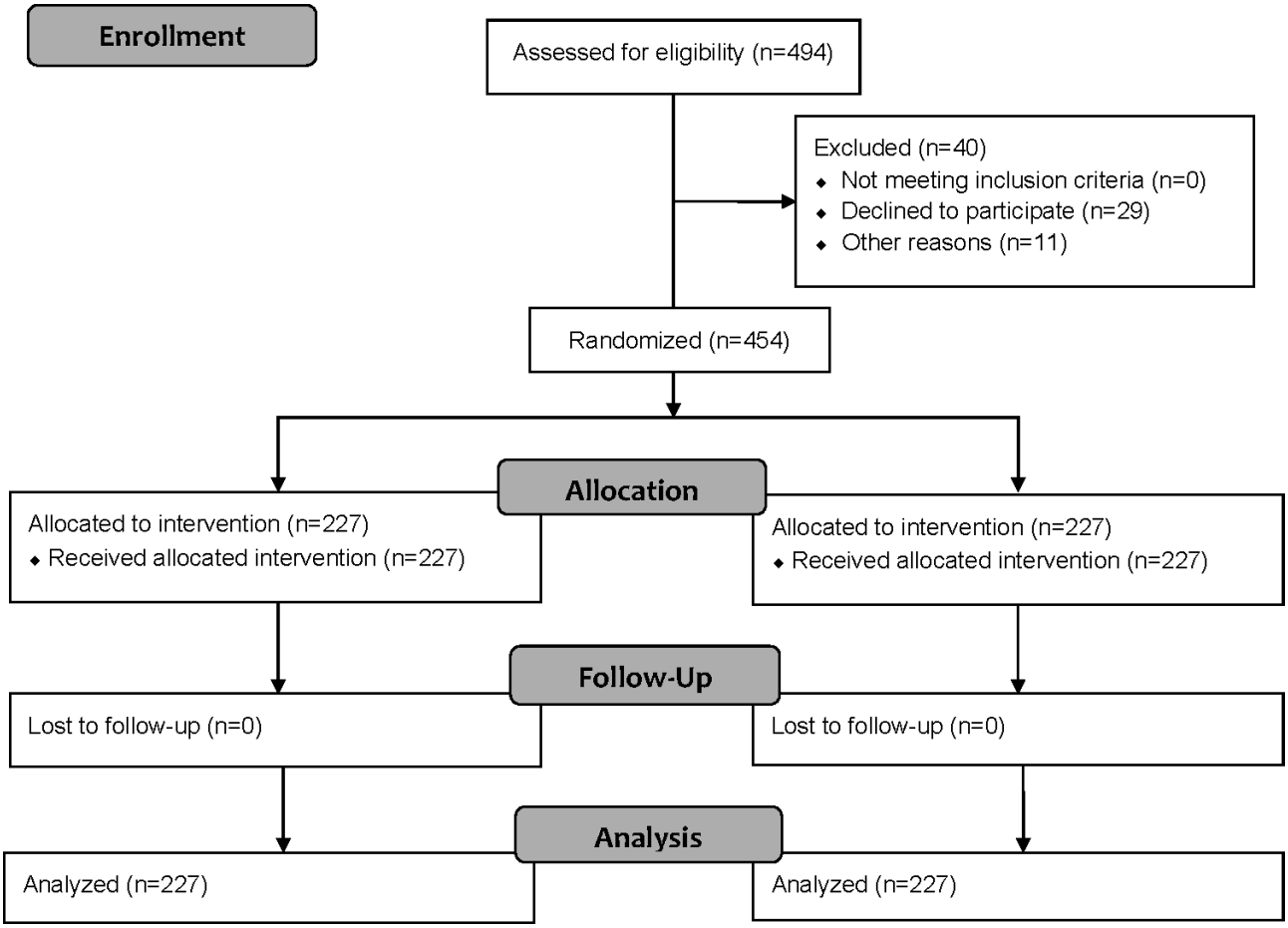
| LGBT (n=176) | Non-LGBT (n=278) |
|---|---|
| Liberal (n=84, 47.7%) | Liberal (n=103, 37.0%) |
| Moderate (n=47, 26.7%) | Moderate (n=63, 22.7%) |
| Conservative (n=45, 25.6%) | Conservative (n=112, 40.3%) |






XSL•FO
RenderX

**Figure 2.** CONSORT (Consolidated Standards of Reporting Trials) 2010 diagram. Checklist is available as Checklist 1.




Participants were 41% (n=187) liberal, 24% (n=110) moderate, and 35% (n=157) conservative; 39% (n=176) were LGBTQ+, with 18% (n=32) identifying as transgender and/or nonbinary. These numbers do not match the intended quotas because individuals' self-identification in terms of politics and/or sexual orientation and gender identity may have changed since they filled out the screeners on the survey platform. (Table 2; Figure 1)



XSL•FO
RenderX

**Table .** Demographics of study participants, by explicit preference for gender-inclusive versus gender-specific information.[a]

| Demographics | Always gender-specific (n=184), n (%) | Sometimes gender-specific (n=59), n (%) | No preference (n=131), n (%) | Sometimes gender-inclusive (n=39), n (%) | Always gender-inclusive (n=41), n (%) | Chi-square (*df*); *P* value |
|---|---|---|---|---|---|---|
| How old are you? (y) | | | | | | 28.9 (16); .02[b] |
| 18 - 24 | 29 (15.8) | 12 (20.3) | 11 (8.4) | 7 (17.9) | 6 (14.6) | |
| 25 - 34 | 48 (26.1) | 16 (27.1) | 49 (37.4) | 16 (41.0) | 15 (36.6) | |
| 35 - 44 | 43 (23.4) | 15 (25.4) | 37 (28.2) | 8 (20.5) | 13 (31.7) | |
| 45 - 54 | 32 (17.4) | 14 (23.7) | 25 (19.1) | 4 (10.3) | 3 (7.3) | |
| 55 - 64 | 32 (17.4) | 2 (3.4) | 9 (6.9) | 4 (10.3) | 4 (9.8) | |
| Which of the following best describes your racial/ethnic background? | | | | | | 33.7 (28); .21 |
| White | 112 (60.9) | 40 (67.8) | 96 (73.3) | 23 (59.0) | 35 (85.4) | |
| Black | 46 (25.0) | 13 (22.0) | 16 (12.2) | 13 (33.3) | 4 (9.8) | |
| Hispanic | 7 (3.8) | 2 (3.4) | 10 (7.6) | 0 (0) | 1 (2.4) | |
| Asian or Pacific Islander | 6 (3.3) | 2 (3.4) | 1 (0.8) | 2 (5.1) | 0 (0) | |
| Native American | 2 (1.1) | 0 (0) | 1 (0.8) | 0 (0) | 1 (2.4) | |
| Multiracial | 9 (4.9) | 2 (3.4) | 7 (5.3) | 1 (2.6) | 0 (0) | |
| Something else | 1 (0.5) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | |
| Prefer not to answer | 1 (0.5) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | |
| Were you born in the United States? | | | | | | 10.2 (8); .25 |
| Yes, and so were my parents | 151 (82.1) | 49 (83.1) | 115 (87.8) | 31 (79.5) | 39 (95.1) | |
| Yes, but my parents were not | 21 (11.4) | 4 (6.8) | 10 (7.6) | 3 (7.7) | 1 (2.4) | |
| No | 12 (6.5) | 6 (10.2) | 6 (4.6) | 5 (12.8) | 1 (2.4) | |
| Highest level of education achieved | | | | | | 47.1 (28); .01[b] |
| Less than high school | 0 (0) | 0 (0) | 2 (1.5) | 0 (0) | 2 (4.9) | |
| High school diploma or General Educational Development (GED) test | 17 (9.2) | 6 (10.2) | 29 (22.1) | 4 (10.3) | 2 (4.9) | |
| Some college | 29 (15.8) | 10 (16.9) | 24 (18.3) | 5 (12.8) | 10 (24.4) | |
| Associate's degree | 22 (12.0) | 7 (11.9) | 15 (11.5) | 2 (5.1) | 9 (22.0) | |
| Bachelor's degree | 79 (42.9) | 20 (33.9) | 39 (29.8) | 17 (43.6) | 11 (26.8) | |
| Master's degree | 29 (15.8) | 14 (23.7) | 19 (14.5) | 11 (28.2) | 5 (12.2) | |
| Professional degree | 2 (1.1) | 0 (0.0) | 2 (1.5) | 0 (0) | 1 (2.4) | |
| Doctorate degree | 6 (3.3) | 2 (3.4) | 1 (0.8) | 0 (0) | 1 (2.4) | |
| Which of the following best describes your political orientation? | | | | | | 154.3 (16); <.001[b] |



XSL•FO
RenderX

| Demographics | Always gender-specific (n=184), n (%) | Sometimes gender-specific (n=59), n (%) | No preference (n=131), n (%) | Sometimes gender-inclusive (n=39), n (%) | Always gender-inclusive (n=41), n (%) | Chi-square (*df*); *P* value |
|---|---|---|---|---|---|---|
| Very Liberal | 11 (6.0) | 10 (16.9) | 37 (28.2) | 17 (43.6) | 27 (65.9) | |
| Somewhat Liberal | 25 (13.6) | 16 (27.1) | 28 (21.4) | 7 (17.9) | 9 (22.0) | |
| Moderate | 37 (20.1) | 19 (32.2) | 44 (33.6) | 6 (15.4) | 4 (9.8) | |
| Somewhat Conservative | 56 (30.4) | 9 (15.3) | 17 (13.0) | 2 (5.1) | 1 (2.4) | |
| Very Conservative | 55 (29.9) | 5 (8.5) | 5 (3.8) | 7 (17.9) | 0 (0.0) | |
| Which of the following comes closest to how you think about the place of religion in your life? [Religion is…] | | | | | | 96.8 (16); <.001[b] |
| The most important thing | 52 (28.3) | 5 (8.5) | 8 (6.1) | 2 (5.1) | 1 (2.4) | |
| One among many important things | 78 (42.4) | 21 (35.6) | 32 (24.4) | 14 (35.9) | 8 (19.5) | |
| Not as important as other things | 21 (11.4) | 9 (15.3) | 19 (14.5) | 6 (15.4) | 3 (7.3) | |
| Not important | 31 (16.8) | 24 (40.7) | 68 (51.9) | 16 (41.0) | 25 (61.0) | |
| Prefer not to answer | 2 (1.1) | 0 (0) | 4 (3.1) | 1 (2.6) | 4 (9.8) | |
| Do you routinely visit a gynecologist and/or get gynecological/reproductive health care? | | | | | | 16.3 (12); .18 |
| Yes, at least once every two years | 116 (63.0) | 34 (57.6) | 69 (52.7) | 25 (64.1) | 21 (51.2) | |
| Yes, at least once every five years | 29 (15.8) | 6 (10.2) | 22 (16.8) | 4 (10.3) | 6 (14.6) | |
| No, but I have in the past | 28 (15.2) | 16 (27.1) | 35 (26.7) | 8 (20.5) | 8 (19.5) | |
| No, and I prefer to not get this type of care | 11 (6.0) | 3 (5.1) | 5 (3.8) | 2 (5.1) | 6 (14.6) | |
| Do you consider yourself to be LGBTQ+[c] (ie, a sexual or gender minority)? | | | | | | 27.7 (8); <.001[b] |
| No | 132 (71.7) | 31 (52.5) | 70 (53.4) | 23 (59.0) | 16 (39.0) | |
| Yes | 51 (27.7) | 27 (45.8) | 59 (45.0) | 14 (35.9) | 25 (61.0) | |
| I don’t know | 1 (0.5) | 1 (1.7) | 2 (1.5) | 2 (5.1) | 0 (0) | |
| Do you identify as transgender and/or nonbinary? | | | | | | 28.3 (8); <.001[b] |
| No | 44 (86.3) | 23 (85.2) | 51 (87.9) | 8 (57.1) | 12 (48.0) | |
| Yes | 6 (11.8) | 2 (7.4) | 7 (12.1) | 5 (35.7) | 12 (48.0) | |
| I don’t know | 1 (2.0) | 2 (7.4) | 0 (0) | 1 (7.1) | 1 (4.0) | |
| Do you think that sex education should be required as part of public K-12 education? | | | | | | 62.4 (16); <.001[b] |



XSL•FO
RenderX

| Demographics | Always gender-specific (n=184), n (%) | Sometimes gender-specific (n=59), n (%) | No preference (n=131), n (%) | Sometimes gender-inclusive (n=39), n (%) | Always gender-inclusive (n=41), n (%) | Chi-square (*df*); *P* value |
|---|---|---|---|---|---|---|
| Yes, and it should be comprehensive | 100 (54.6) | 45 (76.3) | 94 (72.3) | 31 (79.5) | 41 (100) | |
| Yes, and it should emphasize abstinence until marriage | 19 (10.4) | 2 (3.4) | 10 (7.7) | 2 (5.1) | 0 (0) | |
| Maybe, and if offered it should be comprehensive | 21 (11.5) | 8 (13.6) | 21 (16.2) | 3 (7.7) | 0 (0) | |
| Maybe, and if offered, it should emphasize abstinence until marriage | 25 (13.7) | 2 (3.4) | 2 (1.5) | 3 (7.7) | 0 (0) | |
| Sex education should not be allowed in public schools | 18 (9.8) | 2 (3.4) | 3 (2.3) | 0 (0) | 0 (0) | |

[a]The column headings represent the options for the following item: "When thinking about the health information that you read online or get from a health care provider, do you like it to be:". Differences across these categories were calculated using the chi-square test.

[b]Statistically significant at $P<.05$.

[c]LGBTQ+: lesbian, gay, bisexual, transgender, queer, intersex, and asexual, as well as other sexual and gender minority.

## Participants Stated (Explicit) Preferences Around Gender-Inclusive Language

After individuals were debriefed, and the purpose of the study was explained, they were asked, "When thinking about the health information that you read online or get from a health care provider, would you like it to" and then given the choices of "always be gender-specific (eg, women's health, pregnant women, mother, etc)" (n=184, 40.5%), "sometimes be gender-specific" (n=59, 13%), "no preference either way" (n=131, 28.8%), "sometimes be degendered (eg, reproductive and sexual health, pregnant people, parent, etc)" (n=39, 8.6%), or "always be degendered" (n=41, 9%). We refer to this as their stated or explicit preference for the remainder of the analysis.

Individuals who stated they preferred inclusive language tended to be younger ($\chi^2_{16}=28.9$; $P=.02$) and more liberal ($\chi^{21}_{16}=154.3$; $P<.001$) and also considered religion less important ($\chi^2_{16}=96.8$; $P<.001$) than those who preferred gender-specific language. They were also more likely to be LGBT ($\chi^2_8=27.7$; $P<.001$) and specifically transgender or nonbinary ($\chi^2_8=28.3$; $P<.001$). Most strikingly, the differences in preference were strongly associated with differences in perspectives on sex education in public schools, which was used as a proxy for people's overall beliefs on reproductive health education ($\chi^2_{16}=62.4$; $P<.001$). Every person who preferred inclusive language supported comprehensive sex education in public schools, while no one who thought sex education should not be allowed in public schools had an explicit preference for inclusive language.

## People's Awareness of Inclusive Language

Of the people who saw the inclusive versions first, 56.9% (n=132) stated they did not notice a difference between the documents in the first pair, and 41.4% (n=94) stated they did not notice a difference between the documents in the second pair. These numbers were 61.6% (n=141) and 42.2% (n=96) for those who saw the gender-specific versions first. After the debrief, when asked explicitly about when they noticed that some of the documents were inclusive, 57.7% (n=262) stated that they did not notice until the debrief, 21.1% (n=96) stated that they noticed after the first pair, and 21.1% (n=96) stated that they noticed after the second pair. This did not differ by version ($\chi^2_2=0.6$; $P=.75$). Individuals with the strongest preferences (ie, always) tended to notice sooner than those with weaker (ie, sometimes) or no preference ($\chi^2_8=56.3$; $P<.001$; Table 3).



XSL•FO
RenderX

**Table .** Relationship between stated (explicit) preference for gender-specific versus gender-inclusive content at the time of debrief and measured content preferences.[a]

| Stated preference for pair | Always be gender-specific (n=184) | Sometimes be gender-specific (n=59) | No preference (n=131) | Sometimes be gender-inclusive (n=39) | Always be gender-inclusive (n=41) | Chi-square or ANOVA F (*df*); *P* value |
|---|---|---|---|---|---|---|
| PCOS[b], n (%) | | | | | | 11.2 (3)[c]; .01 |
| Concordant preference[d] | 45 (24.5) | 13 (22.0) | —[e] | 10 (25.6) | 20 (48.8) | |
| Do not notice or discordant preference | 139 (75.5) | 46 (78.0) | — | 29 (74.4) | 21 (51.2) | |
| Birth control, n (%) | | | | | | 9.8 (3)[c]; .02 |
| Concordant preference | 69 (37.5) | 25 (42.4) | — | 10 (25.6) | 24 (58.5) | |
| Do not notice or discordant preference | 115 (62.5) | 34 (57.6) | — | 29 (74.4) | 17 (41.5) | |
| When did you notice that one of each set of the content was gender-inclusive? n (%) | | | | | | 56.3 (8)[c]; <.001 |
| After first | 53 (28.8) | 9 (15.3) | 8 (6.1) | 10 (25.6) | 16 (39.0) | |
| After second | 36 (19.6) | 13 (22.0) | 20 (15.3) | 16 (41.0) | 11 (26.8) | |
| At debrief | 95 (51.6) | 37 (62.7) | 103 (78.6) | 13 (33.3) | 14 (34.1) | |
| Preference scores (higher=stronger preference for gender-specific), mean (SD) | | | | | | |
| Content 1 (PCOS) | 0.6 (2.4) | 0.3 (2.4) | 0.0 (1.1) | 0.3 (1.7) | −1.4 (2.9) | 11.5 (1)[f]; <.001 |
| Content 2 (birth control) | 0.2 (2.3) | −0.1 (2.1) | −0.0 (1.7) | 0.2 (2.2) | −1.7 (3.2) | 8.2 (1)[f]; <.001 |

[a]The column headings represent the options for the following item: “When thinking about the health information that you read online or get from a healthcare provider, do you like it to.”

[b]PCOS: polycystic ovary syndrome.

[c]Chi-square.

[d]For always or sometimes gender-specific, a concordant preference was defined as preferring the gender-specific document. For always or sometimes gender-inclusive, a concordant preference was defined as preferring the gender-inclusive document.

[e]Not applicable, individuals who did not have an explicit preference were excluded from this portion of the analysis.

[f]ANOVA F.

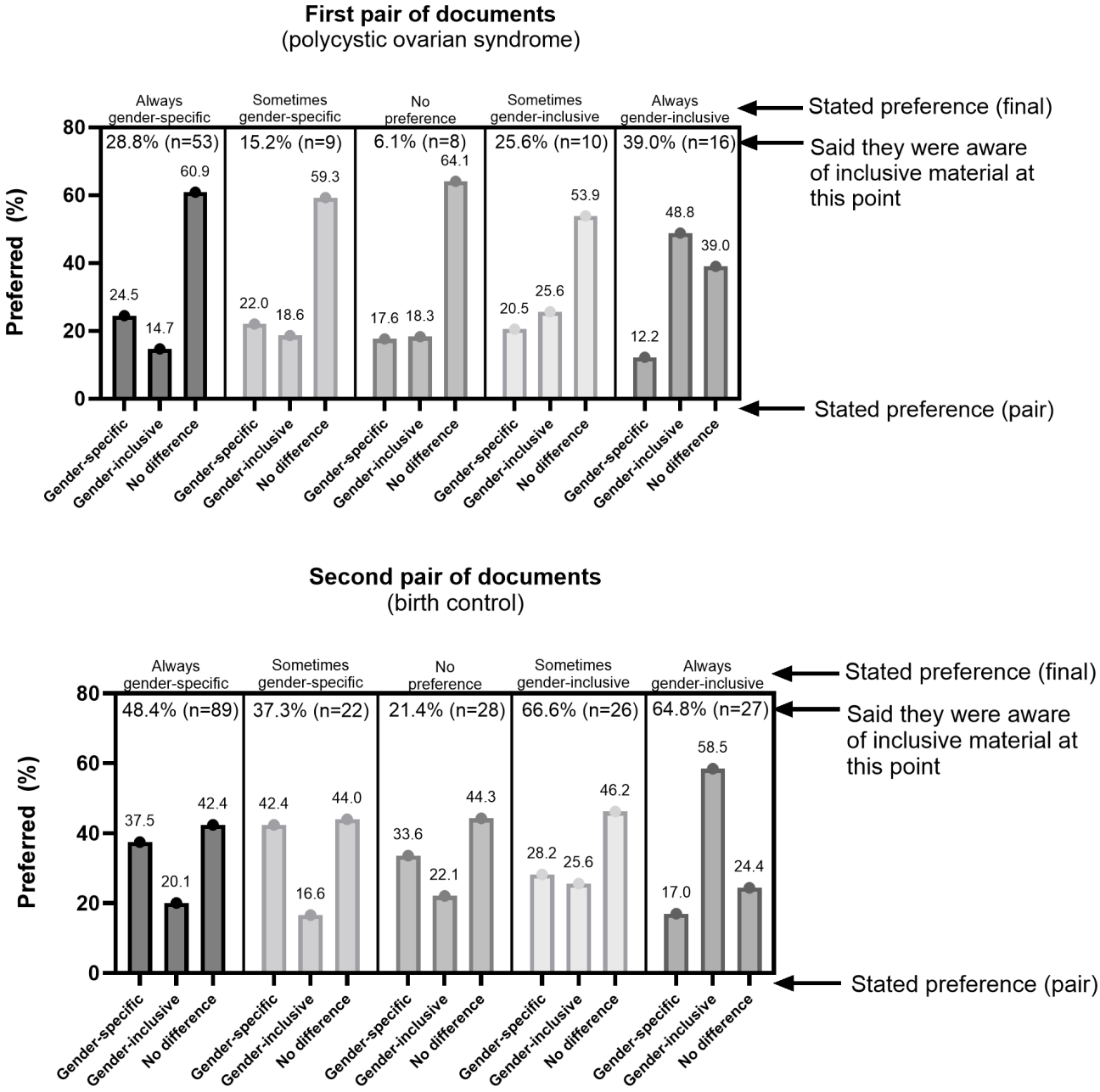
## People’s Version Preferences in Relation to Their Stated (Explicit) Preferences

Looking just at individuals who had a stated preference for gender-specific or gender-inclusive documents (ie, excluding those who chose “no preference”), the majority of individuals who had a stated preference for gender-specific health education documents *did not choose* the gender-specific document as their preferred version for either the first or second pair of documents, although the number of individuals with a concordant preference did increase for the second document pair. The individuals who preferred content to always be inclusive were substantially more likely to choose the concordant version of their document than those from other stated preference groups (n=20, 48.8% for the first pair; n=24, 58.5% for the second pair). This pattern was also evident in the preference scores generated by questions asked after each individual document, where the only group with large differences in rating scores between the paired versions was those who wanted content to always be inclusive (Table 3, Figure 3).



XSL•FO
RenderX

**Figure 3.** Stated preferences for each pair of documents (bars) organized by the final explicit preference stated for gender-specific versus gender-inclusive content (sections). The percentage of individuals who stated, at the end of the study, that they were aware of degendering after viewing the document pair is shown at the top of each section.




## Qualitative Analysis of Open-Ended Responses

We hypothesized during study design and data collection that most participants who preferred gender-specific language would not notice inclusive language, even when they wanted the language to always be gender-specific. While this was true in the quantitative responses, analysis revealed that some people felt strongly about minor differences between the documents. This manifested qualitatively in the free-text responses, which were categorized using inductive content analysis (Multimedia Appendix 1).

When reviewing the comments on why individuals preferred their chosen versions, the category with the greatest number of comments was "perception versus reality," dominated by those who appeared to perceive substantially more differences than were actually present between document versions (n=82), which they used to justify their preferences (ie, citing things such as different layouts, flow, and comprehensiveness). That category also included individuals commenting during the debrief, usually with surprise, that they had not noticed the difference between the pairs of documents (n=21). The next most common category of comment was those related to discourse including comments mentioning inclusiveness (n=33), bioessentialism (n=26), and the use of specific language from media tropes (n=11). Finally, there were groups of comments evaluating specific types of language (gender-inclusive, n=27; gender-specific, n=14; problems with inclusive, n=12; pros and cons of both, n=8), personal preferences for gender-specific (n=14) versus gender-inclusive (n=13) content, and comments mentioning identity or writing quality (Multimedia Appendix 2).



XSL•FO
RenderX

## Discussion

### Principal Findings

Confirming our primary hypothesis, the majority of study participants who were opposed to the use of inclusive language either did not notice a difference between gender-specific and gender-inclusive materials or preferred the materials with inclusive language. Those individuals who had a strong preference for inclusive information were significantly more likely to both notice the differences and choose the concordant document than those with other preferences. Nonetheless, it is important to acknowledge that half of that group still did not notice a difference between the materials when the first pair was presented to them.

The study results also confirmed our secondary hypothesis. Individuals had stronger explicit or stated preferences about gender-inclusive language than were represented by their actual ratings and choices of different types of content. In other words, it appeared that people had opinions about this issue even when the presence of inclusive language did not affect them in practice. This was demonstrated by the fact that what people stated as their explicit preference for gender-specific information did not actually predict the content they said they preferred, but it did predict their political views.

Our data suggest that individuals may hold beliefs about the concept of inclusion that differ from the reality of how they experience it. For some people, this may reflect their having been influenced by media and other forms of social polarization to be upset about the concept of inclusion rather than the reality of it. It is also possible that outside of the echo chamber of highly polarizing virtual communities, the importance or performance of highly gendered and explicitly anti-inclusive preferences is reduced. The act of engaging with potentially applicable health content about the *real self* may support different attention to that content than the *virtual self* who may be more engaged with anti-inclusion rhetoric [27]. That said, there were a minority of participants who did appear to be primed for identifying inclusive content, some of whom reproduced common, inaccurate comments about gender and sex in their open-ended responses.

The fact that most people did not appear to perceive a difference between gender-inclusive and gender-specific content when they had not been primed to look for it suggests that discourse in the United States about the potential harms of inclusive language, often amplified by self-described ultra-conservatives or right-wing groups, does not reflect the lived realities of most Americans. Hyper-politicization of diversity, equity, and inclusion (DEI) can lead to discussions of these concepts in the context of health care inducing stronger emotional reactions than their actual implementation would [21,28,29]. Our research suggests that, in the coming months and years, it may be important to universally adopt inclusive language as a change in nomenclature that promotes accuracy and does not have significant bearing on the way in which critical health information is perceived by the majority of patients, rather than as an explicitly DEI approach.

### Future Directions

In the future, we hope to analyze the use of inclusive language in real-world situations and in populations of younger patients and their caregivers. Our intention is to decouple the perception of inclusive language from DEI efforts and instead focus on it as a way of accurately and nonjudgmentally describing the bodies of individuals receiving care—similar to the use of accurate anatomic terminology. We also intend to test whether inclusive language is received differently when used in combination with gender-specific language versus on its own (eg, *women and other* people with ovaries vs *people with ovaries*). We hypothesize that combination language will have broader acceptability but possibly also lead to stronger negative reactions in a small minority of people at both ends of the political spectrum—who could become primed to notice and respond to the presence of inclusive language. Our next step in doing so will involve focus groups with a diverse range of people—exploring their reactions to these research results to understand their interpretation and the implications for dissemination. We would also like to use this method to examine people's responses to changing sex and gender language in information unrelated to health—for example, information on the benefits of sports participation for people of *all sexes* versus *all genders*. We hypothesize that subject matter, salience, and valence may all play a role in an individual's response.

### Strengths and Limitations

The strengths of this study include a robust study design and a large, diverse respondent sample. The sample also clearly engaged with the research. This could be seen in the extensive comments across open-ended fields.

The main limitation of this study is that it is impossible to rule out rendering differences that may have contributed to the perception of some users that there were substantial differences between the 2 sets of virtually identical documents. The open-ended responses both confirmed and disconfirmed this. Some participants reported that they had believed the 2 documents in each set to be largely identical. Others seemed to believe that the content was completely different, describing different formatting, different lengths, 1 version being much easier to understand, and so forth.

Because the perception of difference among some users was so substantial, care was taken to replicate the study on multiple machines, varying device types (eg, phone, computer), operating systems, and browsers. Across systems, the only identifiable differences between the materials were the minor differences detailed in Table 1.

The study was also, potentially, limited by the sociopolitical environment in which it took place. Participants completed the study in late December 2024, after the election of Donald Trump but before his inauguration. This may have contributed to experiences of polarization among participants. The study was also specifically designed to capture a politically broad range of the US population and cannot be extrapolated to populations outside the United States.

Because of the sampling choice to purposively recruit for political and LGBTQ+ diversity over other forms of diversity,



XSL•FO
RenderX

the sample is not an accurate representation of the full diversity of the United States. For example, looking at the racial and ethnic makeup of the United States, the proportions of the White (67% vs 59%), Black (20% vs 13%), and multiracial (4% vs 2%) populations were higher in this sample than reported in 2022 Census data, while the Hispanic (4% vs 19%) and Asian (2% vs 6%) populations were lower, and that of American Indian or Alaskan Native population (0.8% vs 0.7%) were similar [30].

### Conclusions

Despite substantial conservative discourse about the ways in which gender-inclusive sexual and reproductive information is offensive and harmful to women, the majority of individuals assigned female at birth do not appear to notice when such content has been made inclusive—even when they state an explicit preference for gender-specific content. The exception is those individuals for whom inclusive content is explicitly relevant or preferred, who both notice and choose content that is concordant with their preferences at much higher rates than those who prefer content to be gender-specific. This suggests that, by and large, inclusive health education is broadly useful and acceptable.

### Acknowledgments

No generative artificial intelligence was used in the writing of this paper or the underlying research.

### Funding

This work was funded by a grant from the Lesbian Health Foundation, part of GLMA Health Professionals Advancing LGBTQ+ Equality.

### Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request, with the completion of a signed data use agreement.

### Authors' Contributions

Conceptualization: ERB, FWG
Formal analysis: ERB, JDK
Funding Acquisition: ERB
Investigation: ERB
Methodology: ERB, JDK
Supervision: FWG
Writing – original draft: ERB, JDK, AKB
Writing – review and editing: ERB, JDK, AKB, FWG

### Conflicts of Interest

None declared.

Multimedia Appendix 1
Major categories and codes derived from inductive content analysis of open-ended responses provided by participants.
[DOCX File, 16 KB - i-jmr_v15i1e85868_app1.docx ]

Multimedia Appendix 2
Content analysis of open-ended responses provided by participants.
[DOCX File, 18 KB - i-jmr_v15i1e85868_app2.docx ]

Checklist 1
CONSORT checklist.
[PDF File, 2841 KB - i-jmr_v15i1e85868_app3.pdf ]

### References

1. Baker AM, Jahn JL, Tan ASL, et al. Sexual health information sources, needs, and preferences of young adult sexual minority cisgender women and non-binary individuals assigned female at birth. Sex Res Social Policy 2021 Sep;18(3):775-787. [doi: 10.1007/s13178-020-00501-6] [Medline: 34484461]
2. McKay EA, Fontenot HB. Parent-adolescent sex communication with sexual and gender minority youth: an integrated review. J Pediatr Health Care 2020;34(5):e37-e48. [doi: 10.1016/j.pedhc.2020.04.004] [Medline: 32861429]



XSL•FO
RenderX

3. Petrizzo K, Moxie J. Past experiences and preferences for LGBTQ+ sex education among LGBTQ+ college students. J Am Coll Health 2025 May;73(5):2010-2020. [doi: 10.1080/07448481.2024.2317173] [Medline: 38442354]
4. Jolly D, Boskey ER, Thomson KA, Tabaac AR, Burns MTS, Katz-Wise SL. Why are you asking? Sexual orientation and gender identity assessment in clinical care. J Adolesc Health 2021 Dec;69(6):891-893. [doi: 10.1016/j.jadohealth.2021.08.015] [Medline: 34629230]
5. Grant R, Nash M, Hansen E. What does inclusive sexual and reproductive healthcare look like for bisexual, pansexual and queer women? Findings from an exploratory study from Tasmania, Australia. Cult Health Sex 2020 Mar;22(3):247-260. [doi: 10.1080/13691058.2019.1584334] [Medline: 30958103]
6. Baldwin A, Dodge B, Schick VR, et al. Transgender and genderqueer individuals' experiences with health care providers: what's working, what's not, and where do we go from here? J Health Care Poor Underserved 2018;29(4):1300-1318. [doi: 10.1353/hpu.2018.0097] [Medline: 30449748]
7. Woolweaver AB, Drescher A, Medina C, Espelage DL. Leveraging comprehensive sexuality education as a tool for knowledge, equity, and inclusion. J Sch Health 2023 Apr;93(4):340-348. [doi: 10.1111/josh.13276] [Medline: 36404494]
8. Blaubergs MS. An analysis of classic arguments against changing sexist language. Womens Stud Int Q 1980 Jan;3(2-3):135-147. [doi: 10.1016/S0148-0685(80)92071-0]
9. Parks JB, Roberton MA. Influence of age, gender, and context on attitudes toward sexist/nonsexist language: is sport a special case? Sex Roles 1998 Mar;38(5-6):477-494. [doi: 10.1023/A:1018766023667]
10. Saguy AC, Williams JA. A little word that means a lot: a reassessment of singular they in a new era of gender politics. Gend Soc 2022 Feb;36(1):5-31. [doi: 10.1177/08912432211057921]
11. Herman JL, Flores AR, O'Niel KK. How many adults and youth identify as transgender in the United States. : The Williams Institute; 2022 URL: https://williamsinstitute.law.ucla.edu/wp-content/uploads/Trans-Pop-Update-Jun-2022.pdf [accessed 2026-06-04]
12. Alstott A, Olgun M, Robinson H, McNamara M. "Demons and Imps": misinformation and religious pseudoscience in state anti-transgender laws. Yale J Law Fem 2024 [FREE Full text]
13. Atwood S, Morgenroth T, Olson KR. Gender essentialism and benevolent sexism in anti‐trans rhetoric. Soc Issues Policy Rev 2024 Jan;18(1):171-193. [doi: 10.1111/sipr.12099]
14. McLean C. The growth of the anti-transgender movement in the United Kingdom. The silent radicalization of the British electorate. Int J Sociol 2021 Nov 2;51(6):473-482. [doi: 10.1080/00207659.2021.1939946]
15. Knauer NJ. LGBTQ rights in an age of democratic uncertainty: how to counter the politics of eradication. SSRN. Preprint posted online on Mar 19, 2025. [doi: 10.2139/ssrn.5161951]
16. Mar SA, Indig G, McNamara M, et al. Intersex exceptions in U.S. laws restricting medical care for transgender minors. medRxiv. Preprint posted online on Apr 22, 2025. [doi: 10.1101/2025.04.22.25326142]
17. Medina C, Babbs G, Mar S, Hughto JMW, Tobin-Tyler E. What do the Trump executive orders mean for the health of trans Americans? Health Affairs Forefront 2025. [doi: 10.1377/forefront.20250213.235899]
18. Gribble KD, Bewley S, Bartick MC, et al. Effective communication about pregnancy, birth, lactation, breastfeeding and newborn care: the importance of sexed language. Front Glob Womens Health 2022;3:818856. [doi: 10.3389/fgwh.2022.818856] [Medline: 35224545]
19. Green E. The culture war over 'pregnant people'. The Atlantic. 2021. URL: https://www.theatlantic.com/politics/archive/2021/09/pregnant-people-gender-identity/620031/ [accessed 2026-05-11]
20. Moseson H, Zazanis N, Goldberg E, et al. The imperative for transgender and gender nonbinary inclusion: beyond women's health. Obstet Gynecol 2020 May;135(5):1059-1068. [doi: 10.1097/AOG.0000000000003816] [Medline: 32282602]
21. Bartick M, Dahlen H, Gamble J, Walker S, Mathisen R, Gribble K. Reconsidering "inclusive language:" consequences for healthcare and equitableness of a growing linguistic movement to address gender identity with a path forward. Sex Reprod Healthc 2025 Jun;44:101088. [doi: 10.1016/j.srhc.2025.101088] [Medline: 40147360]
22. Prolific. URL: https://www.prolific.com/ [accessed 2026-06-05]
23. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
24. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
25. Kyngäs H. Inductive content analysis. In: The Application of Content Analysis in Nursing Science Research: Springer International Publishing; 2019:13-21. [doi: 10.1007/978-3-030-30199-6_2]
26. Croucher SM, Cronn-Mills D. Content analysis. In: Understanding Communication Research Methods, 4th edition: Routledge; 2024:145-159 URL: https://www.taylorfrancis.com/chapters/edit/10.4324/9781003432173-14/content-analysis-stephen-croucher-daniel-cronn-mills [accessed 2026-06-04]
27. Xing Y, Zhang JZ, Storey VC, Koohang A. Diving into the divide: a systematic review of cognitive bias-based polarization on social media. J Enterp Inf Manag 2024 Feb 21;37(1):259-287. [doi: 10.1108/JEIM-09-2023-0459]
28. Smith MA, Myers CL, Chong M, et al. Rooted in white identity politics: tracing the genealogy of critical race theory discourse in identity-based disinformation. Polit Commun 2025 Sep 3;42(5):815-837. [doi: 10.1080/10584609.2025.2458069]




XSL•FO
RenderX

29. Shen MJ, LaBouff JP. More than political ideology: subtle racial prejudice as a predictor of opposition to universal health care among U.S. citizens. J Soc Polit Psych 2016 Aug 18;4(2):493-520. [doi: 10.5964/jspp.v4i2.245]
30. How many people live in the US. USAFacts. 2025. URL: https://usafacts.org/data/topics/people-society/population-and-demographics/our-changing-population/ [accessed 2025-02-15]

## Abbreviations

**CONSORT:** Consolidated Standards of Reporting Trials
**DEI:** diversity, equity, and inclusion
**LGBTQ+:** lesbian, gay, bisexual, transgender, queer, intersex, and asexual, as well as other sexual and gender minority

*Edited by M Balcarras; submitted 14.Oct.2025; peer-reviewed by M Mullins, N Levitz; revised version received 21.May.2026; accepted 22.May.2026; published 22.Jun.2026.*

*Please cite as:*
*Boskey ER, Kant JD, Berman AK, Grimstad FW*
*Determining the Relationship Between People's Explicit and Implicit Preferences for Gender-Inclusive Sexual and Reproductive Health Content: Randomized Controlled Trial*
*Interact J Med Res 2026;15:e85868*
*URL: https://www.i-jmr.org/2026/1/e85868*
*doi:10.2196/85868*

© Elizabeth R Boskey, Jessica D Kant, Ariel K Berman, Frances W Grimstad. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 22.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Association Between Depressive Symptoms and Incidence of Stroke in a Population With Cardiovascular-Kidney-Metabolic Syndrome Stages 0 to 3: Nationwide Prospective Cohort Study

Yunjie Li[1], MB; Lin Wang[2], MM; Lujing Gao[3], MM

[1]Department of Radiology, Zhuzhou Clinical College, Jishou University, Zhuzhou, China

[2]Department of Respiratory and Critical Care Medicine, Zhuzhou Hospital Affiliated to Xiangya School of Medicine, Central South University, Zhuzhou, China

[3]Department of Radiology, Zhuzhou Hospital Affiliated to Xiangya School of Medicine, Central South University, Zhuzhou, China

**Corresponding Author:**
Lujing Gao, MM
Department of Radiology, Zhuzhou Hospital Affiliated to Xiangya School of Medicine, Central South University, Zhuzhou, China

## Abstract

**Background:** The association between depressive symptoms and cardiovascular diseases is well established. However, their impact on the incidence of stroke in individuals with cardiovascular-kidney-metabolic (CKM) syndrome remains unclear.

**Objective:** This study aims to investigate the impact of depressive symptoms at different stages of CKM syndrome on the incidence of new-onset stroke.

**Methods:** This study used data from the China Health and Retirement Longitudinal Study. Depressive symptoms at baseline were assessed using the Center for Epidemiologic Studies Depression Scale, with stroke incidence determined through standardized follow-up questionnaires. Cox regression and restricted cubic spline regression were used to evaluate the association between depressive symptoms and stroke risk.

**Results:** The analysis included 9593 participants (n=5180, 54.92% male; mean age of 60.89, SD 9.39 y), classified into CKM stages 0 to 3. Fully adjusted Cox regression showed that each 1-point increase in depressive score was associated with a 3% higher stroke risk (hazard ratio 1.03, 95% CI 1.02 - 1.04; *P*<.001). Restricted cubic spline regression confirmed a significant positive linear relationship between depressive symptoms and stroke incidence (*P*<.001).

**Conclusions:** This cohort study demonstrates a positive linear association between depressive symptoms and increased stroke incidence in individuals with CKM syndrome (stages 0 - 3). These findings highlight the importance of emotional health management, suggesting that effective depression treatment may help reduce stroke risk through inflammation reduction and lifestyle improvements.

***(Interact J Med Res 2026;15:e80988)*** doi:10.2196/80988

**KEYWORDS**

stroke; cardiovascular-kidney-metabolic syndrome; CKM; depression; restricted cubic spline analysis; RCS; subgroup analysis

## Introduction

In the October 2023 Presidential Advisory, the American Heart Association (AHA) defined cardiovascular-kidney-metabolic (CKM) syndrome as a systemic condition characterized by complex pathophysiological interactions among metabolic risk factors (eg, obesity and diabetes), chronic kidney disease (CKD), and cardiovascular disease (CVD). These interrelated processes contribute to multiorgan dysfunction and significantly increase the risk of adverse cardiovascular events. Investigating CKM syndrome as an integrated condition offers a strong scientific foundation for developing comprehensive diagnostic and therapeutic strategies aimed at improving patient outcomes [1]. Metabolic disorders, by reducing insulin sensitivity, elevate the risk of dyslipidemia in patients with CKM, often leading to early cardiovascular complications such as coronary atherosclerosis [2,3], myocardial infarction [4], and heart failure [5]. Additionally, the interaction between CKD and CVD can trigger inflammatory responses [6], oxidative stress [7], and endothelial dysfunction [8]. Moreover, extensive evidence indicates that CVD constitutes the primary clinical burden associated with CKM syndrome, with a disproportionately high impact [9]. Compared to healthy individuals, patients with CKM experience more complex pathophysiological changes.

Conventional medical practice often diagnoses and treats CVD, CKD, and metabolic disorders separately, overlooking their



XSL•FO
RenderX

interrelationships. AHA emphasizes the importance of early detection of CKM staging in the preclinical phase [1]. Previous studies have shown that adults in the United States have concurrent cardiovascular, kidney, and metabolic diseases, with this prevalence on the rise [10]. Roth et al [11] observed that in patients undergoing noncardiac surgery, the risk of postoperative adverse outcomes increases with higher CKM staging. Certain pharmacological interventions, such as angiotensin-converting enzyme inhibitors or angiotensin II receptor blockers, sodium-glucose co-transporter 2 inhibitors, and nonsteroidal mineralocorticoid receptor antagonists, not only slow the progression of CKD but also reduce mortality from CVD [12]. These findings suggest that the 2 conditions may share common biological pathways. A growing body of evidence suggests that the disproportionate clinical burden of CKM syndrome is strongly linked to CVD [1,13]. Stroke is the second-leading cause of death and the third-leading cause of disability globally [14]. Additionally, it is estimated that 2 million new cases of stroke occur annually in China, with 70% to 80% of affected individuals experiencing disabilities that render them incapable of independent living [15]. Studying the incidence of stroke in populations at different stages of CKM syndrome holds significant clinical and public health value.

In 2013, depression accounted for more than 10 million disability-adjusted life years in China, which is projected to rise by approximately 10% by 2025 [16]. Negative emotional states, such as anxiety, depression, or chronic stress, trigger hyperactivation of the hypothalamic-pituitary-adrenal axis, resulting in elevated cortisol levels that exacerbate metabolic imbalances [17]. This physiological disturbance can lead to increased blood pressure, endothelial dysfunction, and the development of atherosclerosis [18]. Moreover, individuals with depression often exhibit systemic low-grade inflammation, which impairs insulin signaling and contributes to dysregulated glucose metabolism [19]. This condition further promotes vascular endothelial inflammation and oxidative stress, accelerating vascular stiffness and plaque formation [20]. Previous studies have shown a strong association between depression and elevated proinflammatory cytokines (eg, interleukin-6 and C-reactive protein [CRP]) in patients with CKD, potentially increasing the risk of cardiovascular events and overall mortality [21]. A 2-year observational study reported that persistent depressive symptoms elevate the risk of stroke, while improvements in these symptoms reduce its incidence [22]. However, the relationship between depressive symptoms and stroke incidence in the population with CKM remains inadequately explored.

Clarifying the association between depressive symptoms and stroke incidence in this group could inform targeted psychological interventions and disease management strategies, ultimately improving patient outcomes and overall health. This study aims to investigate the impact of depressive symptoms at different stages of CKM syndrome on the incidence of new-onset stroke. Since clinical criteria for CVD are present in CKM stage 4, we selected individuals in CKM stages 0 to 3 as the target population for this study.

## Methods

### Ethical Considerations

The studies involving human participants adhered to the principles of the Declaration of Helsinki and were reviewed and approved by the ethics review committee of Peking University (IRB00001052-11015). The participants provided their written informed consent to participate in this study. To ensure privacy and confidentiality, all data accessed from the China Health and Retirement Longitudinal Study (CHARLS) database were deidentified and analyzed anonymously.

### Data Source and Study Population

The data for this study were obtained from the CHARLS, a national cohort study covering urban and rural areas across 28 provinces in China. CHARLS is representative of the population aged 45 years and older in China. All participants provided written informed consent before participating in the study. The baseline survey for CHARLS began in 2011, including approximately 17,000 respondents from over 10,000 households. A stratified sampling method was used to ensure national representativeness. Follow-up surveys were conducted in 2013, 2015, and 2018. All field staff underwent systematic and professional training and conducted face-to-face interviews using standardized questionnaires. The CHARLS database includes information across 5 key domains: demographic and household characteristics, health status, economic status, health care use and expenditures, and social participation and psychological well-being. In this study, the baseline was set in 2011, with the follow-up period ending in 2018.

### Exposure

The Center for Epidemiologic Studies Depression Scale is a widely used self-report instrument designed to assess depressive symptoms in the general population [23]. The scale contains 10 items measuring depressive symptoms experienced over the past week, including mood states, somatic symptoms, interpersonal difficulties, and positive affect (reverse-scored). Each item is rated on a 4-point Likert scale: 0 (very little or none of the time), 1 (not much; 1 - 2 d), 2 (sometimes or about half the time; 3 - 4 d), and 3 (most of the time; 5 - 7 d). The total score ranges from 0 to 30, with higher scores indicating more severe depressive symptoms. Participants were categorized into four groups based on quartiles of depression scores: the first group ranged from 0 to 3, the second group from 3 to 7, the third group from 7 to 12, and the fourth group from 12 to 30.

### Outcome

The primary outcome of our study was the incidence of new-onset stroke events. A confirmed diagnosis of stroke during the follow-up period was classified as the first event, with the reported date recorded as the onset date. The timing of the stroke was determined by calculating the time interval from the baseline assessment to the reported onset date. For participants who did not report a stroke, the follow-up time was calculated as the interval between the baseline assessment and the date of their last survey. Alternatively, if data loss occurred during this



XSL•FO
RenderX

period, the year in which the loss was first detected was considered the follow-up year.

### Definition of CKM Syndrome Stages 0 to 3

The classification of CKM syndrome stages 0 to 3 is based on the guidelines outlined in the AHA Presidential Advisory Statement on CKM Syndrome [1], while also referencing the CKM classification criteria from the study by Zheng et al [24].

Stage 0 is defined as individuals with no CKM risk factors, characterized by normal BMI, waist circumference, blood glucose levels, blood pressure, and lipid levels, as well as the absence of CKD and CVD.

Stage 1 includes individuals with an elevated BMI (≥23 kg/m² for Asian populations) and an elevated waist circumference (≥80 cm for women; ≥90 cm for men), or prediabetes (defined as a glycated hemoglobin of 5.7% to <6.5% or a fasting blood glucose of 100 mg/dL to <126 mg/dL) without the presence of CKD.

Stage 2 is defined as individuals with metabolic risk factors (eg, hypertriglyceridemia [≥135 mg/dL], hypertension, metabolic syndrome [MetS]) or CKD of moderate-to-high risk and diabetes. MetS is defined by the presence of three or more of the following: (1) ≥80 cm for women and ≥90 cm for men; (2) high-density lipoprotein (HDL) <40 mg/dL for men and <50 mg/dL for women; (3) triglycerides ≥150 mg/dL; (4) elevated blood pressure (systolic blood pressure ≥130 mmHg or diastolic blood pressure ≥80 mmHg and/or use of antihypertensive medications); and (5) fasting blood glucose ≥100 mg/dL.

Stage 3 is characterized by individuals with very-high-risk CKD stages or subclinical atherosclerotic CVD with a predicted 10-year CVD risk of less than 10%. Following the methodology of Deng et al [24], atherosclerotic CVD risk was determined using the Framingham risk score. The estimated glomerular filtration rate (eGFR) was calculated via the Chinese Modification of Diet in Renal Disease equation [25].

### Covariates

We collected data on various independent variables, including demographic characteristics such as age, gender, educational level, BMI, smoking status, alcohol consumption, and place of residence. Blood test parameters such as CRP, uric acid (UA), blood urea nitrogen (BUN), glucose, hemoglobin $A_{1c}$ ($HbA_{1c}$), HDL, non-HDL, low-density lipoprotein (LDL), total cholesterol (TC), triglycerides, eGFR, and platelet count were collected. Additionally, we gathered information on disease conditions, including hypertension, diabetes, CVD, chronic liver disease, sleep disorders, lung disease, and kidney disease. Sleep quality was assessed using the question, “My sleep in the last week was restless,” with 3 response options: little or no time (<1 d), some of the time (1 - 2 d), and occasional or moderate time (3 - 4 d). Participants who answered ≥1 were identified as having a sleep problem. The selection of covariates in this study was based on established theoretical frameworks and prior epidemiological evidence regarding risk factors for both depression [26,27] and CVDs [24].

### Statistical Analysis

Participants in this study were categorized into 4 quartiles (Q1-Q4) based on baseline depressive scores: Q1 (0 - 3), Q2 (4-7), Q3 (8-12), and Q4 (13-30). Continuous variables with a normal distribution were presented as mean (SD), while categorical variables were expressed as frequencies and percentages. Differences between groups were evaluated using the chi-square test and the independent 2-sample 2-tailed *t* test for continuous variables. To explore the association between depressive symptoms and stroke across various demographic characteristics, subgroup analyses were performed for categorical variables such as education level, gender, smoking status, drinking status, and CKM syndrome stages (stages 0-3). The univariate and multivariate Cox regression models were applied to estimate hazard ratio (HR) and 95% CIs for stroke risk associated with baseline depressive scores. Additionally, restricted cubic spline (RCS) regression was used to assess potential nonlinear relationships between depressive scores and stroke incidence in the overall population with CKM syndrome (stages 0 - 3) and specifically within CKM stage 2 and stage 3 subgroups. All statistical analyses were performed using R software (version 4.4.0; R Foundation for Statistical Computing), with a 2-sided *P* value<.05 considered statistically significant.

## Results

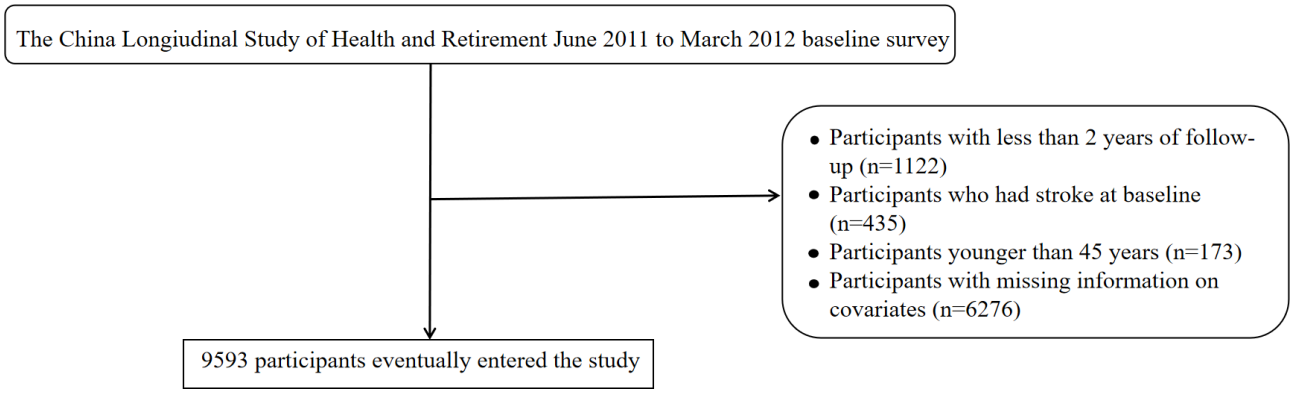
### Baseline Characteristics of Participants

In the 2011 baseline survey, 17,705 participants were initially included. After excluding individuals with missing stroke data, 17,596 participants remained. Further exclusions were made for those with less than 2 years of follow-up (n=1122) and those with a history of stroke at baseline (n=435), leaving a total of 16,039 participants, of whom 12,852 were classified in CKD stages 0 to 3. Subsequently, 173 participants under the age of 45 years were removed. After removing individuals with missing covariate data, 9593 participants were retained for the final analysis (Figure 1).



XSL•FO
RenderX

**Figure 1.** Flowchart illustrating the inclusion and exclusion criteria for the study population.




Among them, 54.92% (n=5268) were male individuals. Baseline characteristics, as shown in Table 1, are presented according to the quartiles of the depressive score. The analysis revealed that women, married individuals, and those with lower educational attainment were more susceptible to depression compared to men, unmarried individuals, and those with higher levels of education. Specifically, there was a stark contrast in the distribution of education levels across depression quartiles: individuals with primary school education or below accounted for 82.18% of the most severe depressive group (Q4), whereas those with a college degree or higher represented only 0.95% of the same group. As depressive symptoms increased, the proportion of women also increased, while the proportion of men decreased ($\chi^2_3$=186.7; *P*<.001). Similarly, the proportion of individuals with lower educational attainment rose, while the proportion of those with higher education decreased ($\chi^2_3$=186.70; *P*<.001), with these differences being statistically significant. Furthermore, individuals with chronic conditions such as CVD, liver disease, kidney disease, and lung disease were more likely to experience depressive symptoms than those without these conditions. Additionally, these individuals were more frequently represented among those with the most severe depressive symptoms (Q4). Specifically, individuals with newly diagnosed stroke exhibited significantly more severe depressive symptoms (mean 9.49, SD 6.53) at the 2011 baseline compared to those without stroke (mean 8.16, SD 6.17). Patients with newly diagnosed stroke had significantly higher levels of age, BMI, LDL, HDL, glucose, triglycerides, TC, UA, CRP, and $HbA_{1c}$ compared to healthy controls. In contrast, their sleep duration and eGFR were significantly lower in patients newly diagnosed with stroke compared to healthy controls (Table 2).



XSL•FO
RenderX

**Table .** Baseline characteristics of individuals classified by quartiles of the depressive score (N=9593)[a].

| Variable | Total (N=9593) | Q1 (n=2495) | Q2 (n=2572) | Q3 (n=2304) | Q4 (n=2222) | *F* test or chi-square (*df*) | *P* value |
|---|---|---|---|---|---|---|---|
| Age (y), mean (SD) | 60.89 (9.39) | 59.63 (9.09) | 60.72 (9.55) | 61.22 (9.57) | 62.17 (9.15) | 30.34 (3)[b] | <.001 |
| Sex, n (%) | | | | | | 186.70 (3)[c] | <.001 |
| Female | 4325 (45.08) | 916 (36.71) | 1072 (41.68) | 1105 (47.96) | 1232 (55.45) | | |
| Male | 5268 (54.92) | 1579 (63.29) | 1500 (58.32) | 1199 (52.04) | 990 (44.55) | | |
| Marital status, n (%) | | | | | | 115.47 (3)[c] | <.001 |
| Married | 8433 (87.91) | 2282 (91.46) | 2314 (89.97) | 2017 (87.54) | 1820 (81.91) | | |
| Unmarried | 1160 (12.09) | 213 (8.54) | 258 (10.03) | 287 (12.46) | 402 (18.09) | | |
| Education, n (%) | | | | | | 350.22 (9)[c] | <.001 |
| College school and higher | 335 (3.49) | 154 (6.17) | 92 (3.58) | 68 (2.95) | 21 (0.95) | | |
| High school | 598 (6.23) | 227 (9.10) | 192 (7.47) | 108 (4.69) | 71 (3.20) | | |
| Middle school | 1899 (19.80) | 619 (24.81) | 562 (21.85) | 414 (17.97) | 304 (13.68) | | |
| Primary school or below | 6761 (70.48) | 1495 (59.92) | 1726 (67.11) | 1714 (74.39) | 1826 (82.18) | | |
| Location, n (%) | | | | | | 151.49 (3)[c] | <.001 |
| Rural | 6293 (65.60) | 1446 (57.96) | 1632 (63.45) | 1561 (67.75) | 1654 (74.44) | | |
| Urban | 3300 (34.40) | 1049 (42.04) | 940 (36.55) | 743 (32.25) | 568 (25.56) | | |
| BMI (kg/m²), mean (SD) | 23.69 (3.93) | 24.08 (3.88) | 23.90 (3.93) | 23.44 (3.84) | 23.25 (4.03) | 22.95 (3)[b] | <.001 |
| Waist, mean (SD) | 85.24 (12.80) | 86.01 (12.95) | 85.83 (12.94) | 84.66 (12.50) | 84.31 (12.67) | 10.35 3)[b] | <.001 |
| Smoking status, n (%) | | | | | | 52.65 (6)[c] | <.001 |
| Former smoke | 982 (10.24) | 240 (9.62) | 274 (10.65) | 249(10.81) | 219 (9.86) | | |
| Never smoke | 5231 (54.53) | 1249 (50.06) | 1382 (53.73) | 1288 (55.90) | 1312 (59.05) | | |
| Smoke | 3380 (35.23) | 1006 (40.32) | 916 (35.61) | 767 (33.29) | 691 (31.10) | | |
| Drinking status, n (%) | | | | | | 91.61 (3)[c] | <.001 |
| Drink | 3448 (35.94) | 1033 (41.40) | 979 (38.06) | 802 (34.81) | 634 (28.53) | | |
| No drink | 6145 (64.06) | 1462 (58.60) | 1593 (61.94) | 1502 (65.19) | 1588 (71.47) | | |
| Sleep problem, n (%) | | | | | | 6.89 (3)[c] | .08 |
| No | 13 (0.14) | 2 (0.08) | 2 (0.08) | 2 (0.09) | 7 (0.32) | | |
| Yes | 9580 (99.86) | 2493 (99.92) | 2570 (99.92) | 2302 (99.91) | 2215 (99.68) | | |
| Platelets, mean (SD) | 210.88 (75.38) | 206.05 (64.24) | 210.10 (71.52) | 212.59 (86.52) | 215.45 (78.57) | 6.63 (3)[b] | <.001 |
| CRP[d], mean (SD) | 2.69 (6.82) | 2.65 (7.07) | 2.60 (5.95) | 2.97 (8.58) | 2.53 (5.23) | 1.92 (3)[b] | .12 |
| Glucose, mean (SD) | 112.64 (39.48) | 113.86 (40.61) | 111.90 (36.90) | 112.22 (39.67) | 112.59 (40.85) | 1.18 (3)[b] | .31 |
| $HbA_{1c}$[e], mean (SD) | 5.32 (0.87) | 5.30 (0.87) | 5.33 (0.85) | 5.31 (0.81) | 5.35 (0.97) | 1.61 (3)[b] | .19 |
| BUN[f], mean (SD) | 15.89 (4.53) | 15.84 (4.35) | 15.97 (4.55) | 15.96 (4.70) | 15.79 (4.53) | 0.88 (3)[b] | .45 |



XSL•FO
RenderX

| Variable | Total (N=9593) | Q1 (n=2495) | Q2 (n=2572) | Q3 (n=2304) | Q4 (n=2222) | *F* test or chi-square (*df*) | *P* value |
|---|---|---|---|---|---|---|---|
| UA[g], mean (SD) | 4.56 (1.27) | 4.70 (1.28) | 4.64 (1.29) | 4.53 (1.27) | 4.36 (1.19) | 32.5 (3)[b] | <.001 |
| TC[h], mean (SD) | 195.73 (38.44) | 195.65 (37.77) | 195.62 (38.94) | 194.86 (39.26) | 196.87 (37.73) | 1.06 (3)[b] | .37 |
| HDL[i], mean (SD) | 50.59 (15.27) | 49.49 (14.97) | 50.28 (15.05) | 50.88 (15.20) | 51.87 (15.82) | 10.16 (3)[b] | <.001 |
| LDL[j], mean (SD) | 117.11 (35.88) | 117.27 (35.87) | 116.98 (35.83) | 116.39 (36.49) | 117.80 (35.33) | 0.61 (3)[b] | .61 |
| Triglycerides, mean (SD) | 141.23 (104.29) | 145.10 (106.06) | 141.91 (107.22) | 139.07 (98.79) | 138.34 (104.32) | 2.08 (3)[b] | .10 |
| Non-HDL, mean (SD) | 145.15 (39.02) | 146.16 (38.48) | 145.34 (39.98) | 143.98 (39.32) | 145.00 (38.15) | 1.28 (3)[b] | .28 |
| eGFR[k], mean (SD) | 120.08 (29.53) | 120.73 (29.22) | 119.13 (26.83) | 119.93 (32.70) | 120.61 (29.36) | 1.56 (3)[b] | .20 |
| Hypertension, n (%) | | | | | | 27.73 (3)[c] | <.001 |
| No | 6447 (67.21) | 1764 (70.70) | 1744 (67.81) | 1523 (66.10) | 1416 (63.73) | | |
| Yes | 3146 (32.79) | 731 (29.30) | 828 (32.19) | 781 (33.90) | 806 (36.27) | | |
| Diabetes, n (%) | | | | | | 26.37 (3)[c] | <.001 |
| No | 8860 (92.36) | 2351 (94.23) | 2392 (93.00) | 2097 (91.02) | 2020 (90.91) | | |
| Yes | 733 (7.64) | 144 (5.77) | 180 (7.00) | 207 (8.98) | 202 (9.09) | | |
| Lung disease, n (%) | | | | | | 129.58 (3)[c] | <.001 |
| No | 8576 (89.40) | 2343 (93.91) | 2325 (90.40) | 2045 (88.76) | 1863 (83.84) | | |
| Yes | 1017 (10.60) | 152 (6.09) | 247 (9.60) | 259 (11.24) | 359 (16.16) | | |
| Liver disease, n (%) | | | | | | 20.10 (3)[c] | <.001 |
| No | 9296 (96.90) | 2444 (97.96) | 2497 (97.08) | 2228 (96.70) | 2127 (95.72) | | |
| Yes | 297 (3.10) | 51 (2.04) | 75 (2.92) | 76 (3.30) | 95 (4.28) | | |
| CVD[l], n (%) | | | | | | 95.58 (3)[c] | <.001 |
| No | 8518 (88.79) | 2309 (92.55) | 2326 (90.44) | 2014 (87.41) | 1869 (84.11) | | |
| Yes | 1075 (11.21) | 186 (7.45) | 246 (9.56) | 290 (12.59) | 353 (15.89) | | |
| Kidney disease, n (%) | | | | | | 44.41 (3)[c] | <.001 |
| No | 9188 (95.78) | 2418 (96.91) | 2486 (96.66) | 2209 (95.88) | 2075 (93.38) | | |
| Yes | 405 (4.22) | 77 (3.09) | 86 (3.34) | 95 (4.12) | 147 (6.62) | | |
| CKM[m] stage, n (%) | | | | | | 14.63 (9)[c] | .10 |
| 0 | 985 (10.27) | 248 (9.94) | 255 (9.91) | 255 (11.07) | 227(10.22) | | |
| 1 | 1817 (18.94) | 502 (20.12) | 523 (20.33) | 407 (17.66) | 385 (17.33) | | |
| 2 | 3999 (41.69) | 1023 (41.00) | 1038 (40.36) | 980 (42.53) | 958 (43.11) | | |



XSL•FO
RenderX

| Variable | Total (N=9593) | Q1 (n=2495) | Q2 (n=2572) | Q3 (n=2304) | Q4 (n=2222) | *F* test or chi-square (*df*) | *P* value |
|---|---|---|---|---|---|---|---|
| 3 | 2792 (29.10) | 722 (28.94) | 756 (29.39) | 662 (28.73) | 652 (29.34) | | |

[a]Data are presented as the mean (SD), median (IQR), or number (%), as appropriate.

[b]*F* values from 1-way ANOVA.

[c]Chi-square test.

[d]CRP: C-reactive protein.

[e]$HbA_{1c}$: hemoglobin $A_{1c}$.

[f]BUN: blood urea nitrogen.

[g]UA: uric acid.

[h]TC: total cholesterol.

[i]HDL: high-density lipoprotein.

[j]LDL: low-density lipoprotein.

[k]eGFR: estimated glomerular filtration rate.

[l]CVD: cardiovascular disease.

[m]CKM: cardiovascular-kidney-metabolic.



XSL•FO
RenderX

**Table .** Baseline characteristics of individuals classified by onset of stroke (N=9593).

| Variable | Total (N=9593) | No (n=8549) | Yes (n=1044) | Chi-square or *t* test (*df*) | *P* value |
|---|---|---|---|---|---|
| Age (y), mean (SD) | 60.89 (9.39) | 60.73 (9.48) | 62.25 (8.46) | −5.43 (9591)[a] | <.001 |
| Sex, n (%) | | | | 0.64 (1)[b] | .42 |
| Female | 4325 (45.08) | 3867 (45.23) | 458 (43.87) | | |
| Male | 5268 (54.92) | 4682 (54.77) | 586 (56.13) | | |
| Marital status, n (%) | | | | 0.87 (1)[b] | .35 |
| Married | 8433 (87.91) | 7525 (88.02) | 908 (86.97) | | |
| Unmarried | 1160 (12.09) | 1024 (11.98) | 136 (13.03) | | |
| Education, n (%) | | | | 7.04 (3)[b] | .07 |
| College school and higher | 335 (3.49) | 295 (3.45) | 40 (3.83) | | |
| High school | 598 (6.23) | 533 (6.23) | 65 (6.23) | | |
| Middle school | 1899 (19.80) | 1724 (20.17) | 175 (16.76) | | |
| Primary school or below | 6761 (70.48) | 5997 (70.15) | 764 (73.18) | | |
| Location, n (%) | | | | 0.05 (1)[b] | .82 |
| Rural | 6293 (65.60) | 5612 (65.65) | 681 (65.23) | | |
| Urban | 3300 (34.40) | 2937 (34.35) | 363 (34.77) | | |
| BMI, mean (SD) | 23.69 (3.93) | 23.57 (3.91) | 24.66 (3.98) | −8.37 (9591)[a] | <.001 |
| Waist, mean (SD) | 85.24 (12.80) | 84.86 (12.71) | 88.41 (13.04) | −8.32 (9591)[a] | <.001 |
| Smoking status, n (%) | | | | 6.12 (2)[b] | .05 |
| Former smoke | 982 (10.24) | 853 (9.98) | 129 (12.36) | | |
| Never smoke | 5231 (54.53) | 4684 (54.79) | 547 (52.39) | | |
| Smoke | 3380 (35.23) | 3012 (35.23) | 368 (35.25) | | |
| Drinking status, n (%) | | | | 2.86 (1)[b] | .09 |
| Drink | 3448 (35.94) | 3098 (36.24) | 350 (33.52) | | |
| No drink | 6145 (64.06) | 5451 (63.76) | 694 (66.48) | | |
| Sleep problem, n (%) | | | | 0.66 (1)[b] | .41 |
| No | 13 (0.14) | 13 (0.15) | 0（0） | | |
| Yes | 9580 (99.86) | 8536 (99.85) | 1044 (100.00) | | |
| Platelets, mean (SD) | 210.88 (75.38) | 209.54 (71.86) | 221.90 (99.04) | −3.91 (9591)[a] | <.001 |
| CRP[c], mean (SD) | 2.69 (6.82) | 2.59 (6.49) | 3.44 (9.10) | −2.91 (9591)[a] | .008 |
| Glucose, mean (SD) | 112.64 (39.48) | 111.92 (38.29) | 118.58 (47.72) | −4.34 (9591)[a] | <.001 |
| $HbA_{1c}$, mean (SD) | 5.32 (0.87) | 5.30 (0.86) | 5.45 (0.97) | −4.66 (9591)[a] | <.001 |
| BUN[e], mean (SD) | 15.89 (4.53) | 15.89 (4.54) | 15.90 (4.45) | −0.08 (9591)[a] | .94 |
| UA[f], mean (SD) | 4.56 (1.27) | 4.55 (1.26) | 4.70 (1.34) | −3.45 (9591)[a] | <.001 |
| TC[g], mean (SD) | 195.73 (38.44) | 195.32 (38.54) | 199.11 (37.43) | −3.08 (9591)[a] | .006 |
| HDL[h], mean (SD) | 50.59 (15.27) | 50.88 (15.26) | 48.16 (15.16) | 5.46 (9591)[a] | <.001 |
| LDL[i], mean (SD) | 117.11 (35.88) | 116.80 (35.83) | 119.59 (36.26) | −2.35 (9591)[a] | .02 |
| Triglycerides, mean (SD) | 141.23 (104.29) | 139.35 (102.50) | 156.66 (116.83) | −4.58 (9591)[a] | <.001 |



XSL•FO
RenderX

| Variable | Total (N=9593) | No (n=8549) | Yes (n=1044) | Chi-square or *t* test (*df*) | *P* value |
|---|---|---|---|---|---|
| Non-HDL, mean (SD) | 145.15 (39.02) | 144.44 (39.12) | 150.95 (37.69) | −5.25 (9591)[a] | <.001 |
| eGFR[j], mean (SD) | 120.08 (29.53) | 120.63 (29.79) | 115.57 (26.89) | 5.67 (9591)[a] | <.001 |
| Depressive score, mean (SD) | 8.31 (6.23) | 8.16 (6.17) | 9.49 (6.53) | −6.24 (9591)[a] | <.001 |
| Hypertension, n (%) | | | | 232.02 (1)[b] | <.001 |
| No | 6447 (67.21) | 5964 (69.76) | 483 (46.26) | | |
| Yes | 3146 (32.79) | 2585 (30.24) | 561 (53.74) | | |
| Diabetes, n (%) | | | | 54.33 (1)[b] | <.001 |
| No | 8860 (92.36) | 7956 (93.06) | 904 (86.59) | | |
| Yes | 733 (7.64) | 593 (6.94) | 140 (13.41) | | |
| Lung disease, n (%) | | | | 0.20 (1)[b] | .66 |
| No | 8576 (89.40) | 7638 (89.34) | 938 (89.85) | | |
| Yes | 1017 (10.60) | 911 (10.66) | 106 (10.15) | | |
| Liver disease, n (%) | | | | 4.48 (1)[b] | .03 |
| No | 9296 (96.90) | 8296 (97.04) | 1000 (95.79) | | |
| Yes | 297 (3.10) | 253 (2.96) | 44 (4.21) | | |
| CVD[k], n (%) | | | | 70.01 (1)[b] | <.001 |
| No | 8518 (88.79) | 7672 (89.74) | 846 (81.03) | | |
| Yes | 1075 (11.21) | 877 (10.26) | 198 (18.97) | | |
| Kidney disease, n (%) | | | | 17.18 (1)[b] | <.001 |
| No | 9188 (95.78) | 8214 (96.08) | 974 (93.30) | | |
| Yes | 405 (4.22) | 335 (3.92) | 70 (6.70) | | |
| CKM[l] stage, n (%) | | | | 124.24 (3)[b] | <.001 |
| 0 | 985 (10.27) | 945 (11.05) | 40 (3.83) | | |
| 1 | 181(18.94) | 1704 (19.93) | 113 (10.82) | | |
| 2 | 3999 (41.69) | 3488 (40.80) | 511 (48.95) | | |
| 3 | 2792 (29.10) | 2412 (28.21) | 380 (36.40) | | |

[a]*t* test.

[b]Chi-square test.

[c]CRP: C-reactive protein.

[d]$HbA_{1c}$: hemoglobin $HbA_{1c}$.

[e]BUN: blood urea nitrogen.

[f]UA: uric acid.

[g]TC: total cholesterol.

[h]HDL: high-density lipoprotein.

[i]LDL: low-density lipoprotein.

[j]eGFR: estimated glomerular filtration rate.

[k]CVD: cardiovascular disease

[l]CKM: cardiovascular-kidney-metabolic.

The original model examined the relationship between depressive symptoms and the incidence of stroke. In this study, a total of 1044 participants were identified as having experienced a stroke. To evaluate the association between depressive scores and stroke incidence across CKM syndrome stages 0 to 3, 4 Cox proportional hazards models were constructed. Model I was adjusted for age, sex, marital status, educational level, and residential location. Model II expanded upon model I by incorporating additional variables, including smoking status and alcohol consumption. Model III further adjusted for



XSL•FO
RenderX

laboratory parameters such as BUN, platelets, CRP, UA, HDL, LDL, triglycerides, and TC. Model IV extended model III by including comorbid chronic diseases such as hypertension, chronic liver disease, diabetes, and sleep problems. The study found a significant positive association between depression severity and the risk of stroke. When depression severity was treated as a continuous variable, the model yielded an HR of 1.03 (95% CI 1.02 - 1.04; *P*<.001) in the fully adjusted model V.

Further clarifying the relationship between depressive score and stroke incidence, depressive score was categorized into quartiles. In a stratified analysis, the population was divided into 4 groups based on depression severity scores. The second group (scores 3 - 7) showed no significant association with stroke incidence (HR 1.16, 95% CI 0.97 - 1.40; *P*=.10), and this result remained unchanged after adjusting for covariates. In contrast, the third group (HR 1.54, 95% CI 1.29 - 1.84; *P*<.001) and the fourth group (HR 1.69, 95% CI 1.41 - 2.02; *P*<.001) both demonstrated significant associations, with the risk of stroke increasing as depression severity worsened. In the fully adjusted model V, compared to Q1, the HR for Q2, Q3, and Q4 were 1.13 (95% CI 0.94 - 1.35), 1.49 (95% CI 1.24 - 1.79), and 1.63 (95% CI 1.35 - 1.96), respectively.

This indicates that participants in Q2, Q3, and Q4 experienced a 13%, 49%, and 63% higher risk of stroke compared to those in Q1 among the population with CKM syndrome stages 0 to 3. Except for the Q2 stage, the results for the other 2 stages were statistically significant (Table 3).

**Table .** Adjusted hazard ratios (HRs) for stroke incidence across depressive symptom quartiles in a population with cardiovascular-kidney-metabolic (CKM) syndrome stages 0 to 3.

| Character | ScoreQ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Stroke-years, depressive score | Crude model[a] | | Model 1[b] | | Model 2[c] | | Model 3[d] | | Model 4[e] | |
| | HR (95% CI) | *P* value | HR (95% CI) | *P* value | HR (95% CI) | *P* value | HR (95% CI) | *P* value | HR (95% CI) | *P* value |
| Q1 | Reference | | Reference | | Reference | | Reference | | Reference | |
| Q2 | 1.16 (0.97-1.40) | .10 | 1.15 (0.96-1.38) | .13 | 1.15 (0.95-1.38) | .14 | 1.16 (0.97-1.40) | .11 | 1.13 (0.94-1.35) | .21 |
| Q3 | 1.54 (1.29-1.84) | <.001 | 1.54 (1.28-1.84) | <.001 | 1.52 (1.27-1.83) | <.001 | 1.56 (1.30-1.87) | <.001 | 1.49 (1.24-1.79) | <.001 |
| Q4 | 1.69 (1.42-2.02) | <.001 | 1.69 (1.41-2.02) | <.001 | 1.67 (1.39-2.00) | <.001 | 1.73 (1.45-2.08) | <.001 | 1.63 (1.35-1.96) | <.001 |

[a]Crude model: depressive score (Q)

[b]Model 1: depressive scoreQ, adjusted for age, sex, marital status, education, and location. *P* for trend (character2integer)<.001; *P* for trend (median value)<.001.

[c]Model 2: depressive scoreQ adjusted for age, sex, marital status, education, location, smoking status, and drinking status. *P* for trend (character2integer)<.001; *P* for trend (median value)<.001.

[d]Model 3: depressive scoreQ, adjusted for age, sex, marital status, education, location, smoking status, drinking status, blood urea nitrogen (BUN), C-reactive protein (CRP), uric acid (UA), total cholesterol (TC), triglycerides, low-density lipoprotein (LDL), and high-density lipoprotein (HDL). *P* for trend (character2integer)<.001; *P* for trend (median value)<.001.

[e]Model 4: depressive scoreQ, adjusted for age, sex, marital status, education, location, smoking status, drinking status, BUN, CRP, UA, TC, triglycerides, LDL, HDL, hypertension, sleep problem, liver disease, lung disease, and diabetes. *P* for trend (character2integer)<.001; *P* for trend (median value)<.001.

RCS regression models were used to further examine the association between depressive score and the risk of stroke among participants with CKM syndrome. The fully adjusted RCS regression model revealed a significant positive linear association between depression scores and stroke risk across participants with CKM syndrome stages 0 to 3 ($P_{overall}$<.001; $P_{nonlinear}$=.20). Similarly, this positive linear correlation was evident in individuals at CKM syndrome stage 2 ($P_{overall}$<.001; $P_{nonlinear}$=.50) and stage 3 ($P_{overall}$=.03; $P_{nonlinear}$=.30; Figure 2). To further explore the relationship between and the incidence of stroke, subgroup and interaction analyses were conducted across different genders, smoking status, drinking status, chronic diseases, and CKM syndrome stages (stages 0-3) with no interactions observed in other subgroups (*P* for interaction >.05; Table 4 and Figure 3).



XSL•FO
RenderX

**Figure 2.** (A) The restricted cubic spline (RCS) analysis between the depressive score and stroke incidence in a population with cardiovascular-kidney-metabolic (CKM) syndrome stages 0 - 3. The model was adjusted for age, sex, marital status, education, location, smoke, drink, blood urea nitrogen (BUN), C-reactive protein (CRP), uric acid (UA), total cholesterol (TC), triglycerides, low-density lipoprotein (LDL), high-density lipoprotein (HDL), hypertension, sleep problem, liver disease, lung disease, diabetes. (B) The RCS analysis between the depressive score and stroke incidence in a population with CKM syndrome stages 2. (C) The RCS analysis between the depressive score and stroke incidence in a population with CKM syndrome stages 3. HR: hazard ratio.

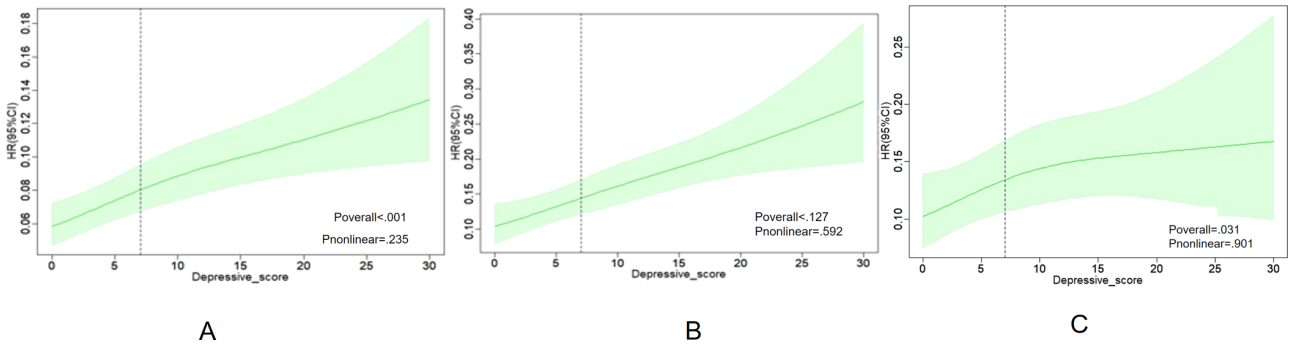


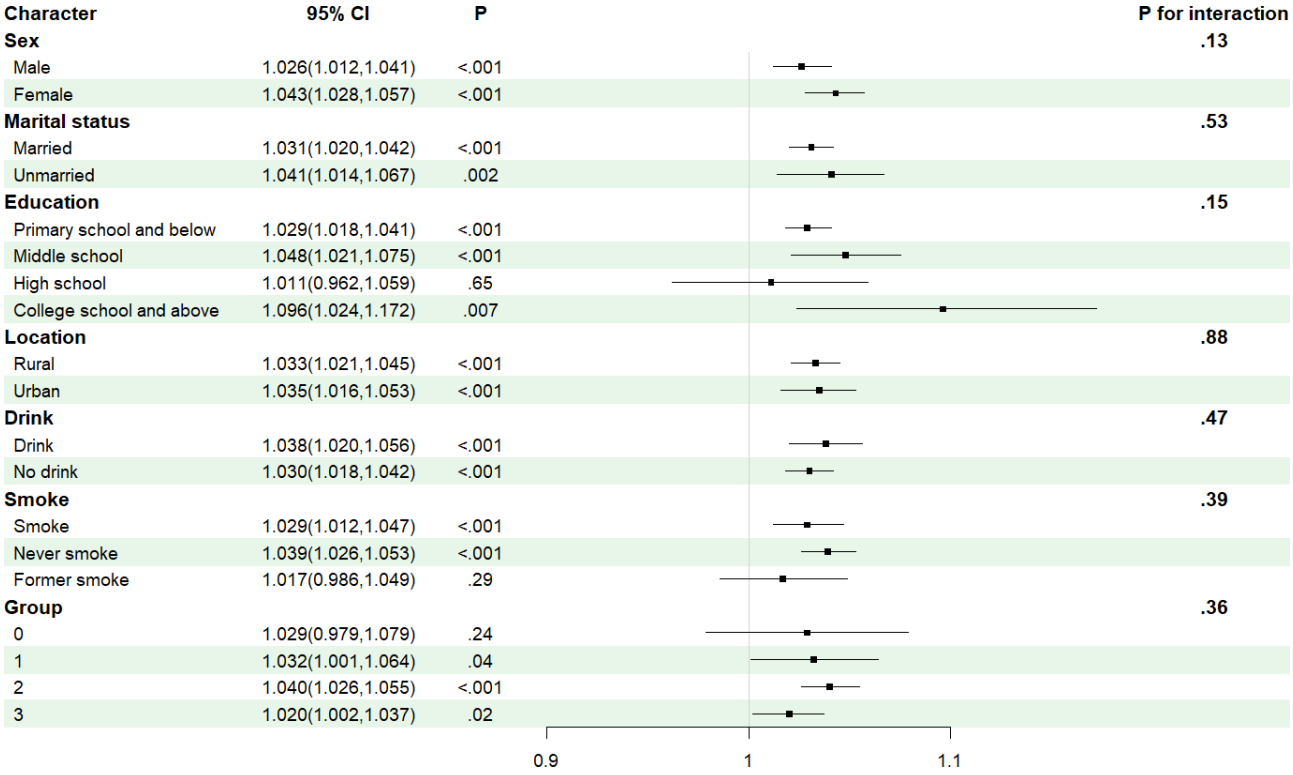

**Table .** Subgroup analyses of the association between the depressive score and stroke incidence in a population with cardiovascular-kidney-metabolic (CKM) syndrome stages 0 to 3.

| Characteristic | HR (95% CI) | *P* value | *P* for interaction |
|---|---|---|---|
| Sex | | | .12 |
| Male | 1.026 (1.012-1.041) | <.001 | |
| Female | 1.043 (1.028-1.057) | <.001 | |
| Marital status | | | .50 |
| Married | 1.031 (1.020-1.042) | <.001 | |
| Unmarried | 1.041 (1.014-1.067) | .002 | |
| Education | | | .15 |
| Primary school or below | 1.029 (1.018-1.041) | <.001 | |
| Middle school | 1.048 (1.021-1.075) | <.001 | |
| High school | 1.011 (0.962-1.059) | .65 | |
| College school and higher | 1.096 (1.024-1.172) | .007 | |
| Location | | | .88 |
| Rural | 1.033 (1.021-1.045) | <.001 | |
| Urban | 1.035 (1.016-1.053) | <.001 | |
| Drinking status | | | .47 |
| Drink | 1.038 (1.020-1.056) | <.001 | |
| No drink | 1.030 (1.018-1.042) | <.001 | |
| Smoke status | | | .39 |
| Smoke | 1.029 (1.012-1.047) | <.001 | |
| Never smoke | 1.039 (1.026-1.053) | <.001 | |
| Former smoke | 1.017 (0.986-1.049) | .29 | |
| Stage | | | .36 |
| 0 | 1.029 (0.979-1.079) | .24 | |
| 3 | 1.020 (1.002-1.037) | .024 | |
| 2 | 1.040 (1.026-1.055) | <.001 | |
| 1 | 1.032 (1.001-1.064) | .04 | |



XSL•FO
RenderX

**Figure 3.** The forest plot of subgroup analyses of the association between the depressive score and stroke incidence in a population with cardiovascular-kidney-metabolic (CKM) syndrome stages 0 to 3.




## Discussion

### Principal Findings

This study extends current knowledge by establishing for the first time a significant association between depressive symptoms and stroke incidence in individuals with CKM syndrome—a condition marked by the intricate interplay of cardiovascular, kidney, and metabolic dysfunctions. Through multivariable-adjusted Cox proportional hazards regression analyses controlling for demographic characteristics (age, sex, and education), lifestyle factors (smoking and physical activity), and clinical parameters (blood pressure, glycemic status, and renal function), we observed a 3% incremental stroke risk per 1-point increase in depressive symptom severity (adjusted HR 1.03, 95% CI 1.02 - 1.04; $P$<.001). Notably, quartile-stratified analyses revealed a striking dose-response gradient. Participants in the highest depressive symptom quartile exhibited 63% greater stroke risk compared to the lowest quartile (HR 1.63, 95% CI 1.35 - 1.96; $P_{trend}$<.001), with intermediate quartiles demonstrating proportionally elevated risks (Q2: HR 1.13, 95% CI 0.94-1.35; Q3: HR 1.49, 95% CI 1.24-1.79). This monotonic risk escalation suggests a potential threshold effect where severe depressive symptoms may constitute a clinically meaningful cerebrovascular risk marker in populations with CKM.

This potential bidirectional relationship highlights the complex interplay between depression and stroke, suggesting that while depressive symptoms may precede stroke, the physiological and psychological impact of a stroke event may also exacerbate depressive states. In depression, disruptions in the serotonin system [28] and hyperactivation of the hypothalamic-pituitary-adrenal axis [29], characteristic of depression, may lead to elevated levels of inflammatory cells and proinflammatory cytokines such as IL-6 and IL-10 [30], both of which are closely associated with an increased risk of stroke. Chronic inflammation accelerates vascular injury and promotes atherosclerosis, significantly elevating the risk of both ischemic and hemorrhagic stroke. Additionally, excessive reactive oxygen species–induced oxidative stress further amplifies this risk by damaging endothelial cells and compromising vascular integrity [31]. Conversely, the inflammatory response following a stroke can deplete L-tryptophan, the precursor of serotonin [32], thereby worsening serotonin system dysfunction and contributing to depressive symptoms such as anhedonia [33]. Regarding the relationship between depression and kidney disease. A systematic review and meta-analysis found that in patients with CKD, depressive symptoms are significantly associated with higher levels of inflammatory markers, suggesting that inflammation may be a shared underlying mechanism linking depression with both CKD and stroke [34]. A study on patients with hypertension found that higher serum cortisol levels are associated with lower eGFR and poorer renal function [35]. Similarly, the pathological mechanisms of the CKM syndrome involve persistent chronic inflammation, oxidative stress, and endocrine dysfunction [36]. Abnormal fat accumulation in CKM serves as a major source of proinflammatory cytokines, exacerbating systemic inflammatory responses and further impairing cardiovascular and renal function [37]. However, this study did not observe a significant positive correlation between the severity of depressive symptoms and stroke incidence in patients with CKM stage 0. Chronic inflammation associated with CKM, along with substantial alterations in individual metabolic or cardiovascular risk factors, may play a pivotal role in mediating this relationship. Further research is warranted to elucidate these



XSL•FO
RenderX

complex interactions and to identify potential intervention targets to mitigate stroke risk in this population.

Our analysis revealed that participants experiencing severe depressive symptoms in the CKM syndrome group had an increased risk of stroke compared to those with less severe symptoms. Stratified subgroup analyses (eg, age and gender) further confirmed that depressive symptoms significantly correlate with new-onset stroke incidence. These findings emphasize that depression is an independent risk factor for stroke, even when other systemic diseases are present. Addressing mild-to-moderate depressive symptoms early to prevent progression could significantly reduce the risk of stroke in populations with CKM syndrome, potentially mitigating both individual and systemic health burdens.

While numerous studies have documented a connection between depression and stroke, these investigations often overlook the broader systemic interactions underlying this relationship. Most previous studies have focused on the impact of poststroke depression on patient outcomes. Sun et al identified that disruptions in neurovascular coupling following a stroke may be a key mechanism underlying the development of depressive symptoms [38]. Approximately 27% of individuals who had a stroke experienced depression, with early-onset depression often persisting [39]. While research on the impact of depression on stroke incidence remains limited, most studies are cross-sectional. Persistently high depressive symptoms and remitted depressive symptoms were associated with increased stroke risk in a large cohort of US adults [40]. A meta-analysis encompassing 28 prospective cohort studies found that depression is associated with a 55% increase in the risk of stroke-related mortality [41]. Using the CHARLS database and logistic regression methods, Cui et al [42] identified a strong association between persistently high levels of depressive symptoms and increased stroke risk. However, the study overlooked the time-dependent relationship between depressive symptoms and stroke onset. In this study, we used a Cox proportional hazards model to examine the association between baseline depressive symptoms and the incidence of stroke, which accounts for time-dependent covariates and is better suited to complex clinical research scenarios.

This study offers several strengths. First, by using data from a large-scale prospective longitudinal study, we used Cox regression models to establish a positive association between depressive symptoms and stroke incidence. Notably, this is the first study to consider the influence of time-dependent variables on outcomes within a syndrome-specific population, thereby improving its clinical relevance. Second, we investigated both the overall relationship between depressive symptoms and stroke risk and the association across different quartiles of depressive symptom severity, allowing for a more nuanced understanding of how varying levels of depression impact stroke occurrence. Finally, subgroup analyses were performed to confirm the robustness and reliability of our findings.

This study has several limitations. First, previous research, including the China Kadoorie Biobank study [43] and other studies on Chinese populations, highlights the low proportion of individuals with depression who receive treatment [44]. As a result, the impact of antidepressant therapy on the study outcomes is likely minimal. Second, the diagnoses of stroke, diabetes, and hypertension in this study were based on self-reported information from consultations with medical professionals, which may introduce reporting bias. To address this limitation, future research should incorporate more objective diagnostic measures. Additionally, depressive symptoms were assessed using a single baseline Center for Epidemiologic Studies Depression Scale score, which may not capture dynamic longitudinal changes during the follow-up period. However, baseline measurements are validated as independent predictors in large-scale cohorts and represent a critical window for early clinical screening and risk stratification. Moreover, the incidence of stroke, as well as diagnoses of hypertension and diabetes, relied on self-reported data, which could lead to recall and misclassification biases. Finally, despite adjusting for a wide range of covariates, residual confounding may still exist due to unmeasured factors, such as specific dietary patterns, physical activity, and the use of medications like antidepressants or antihypertensives. While previous research suggests that antidepressant treatment rates are particularly low in Chinese populations, minimizing their potential impact on our findings, future studies should integrate more objective diagnostic measures and longitudinal data to further strengthen the robustness of these associations.

## Conclusions

This cohort study provides evidence of a significant association between depressive symptoms and an increased risk of stroke in individuals with CKM syndrome stages 0 to 3. The new AHA definition underscores the impact of emotional management on systemic, multisystem diseases. Effective treatment of depression may indirectly reduce the risk of stroke by lowering inflammation levels and promoting healthier lifestyle changes.

## Acknowledgments

The authors confirm that no generative artificial intelligence (AI) or AI-assisted technologies were used in the generation of any part of this manuscript, including the research design, data analysis, or text writing.

## Authors' Contributions

YL and LW designed the study and analyzed the data. YL and LW performed literature research and created the tables and plots. YL and LW drafted the manuscript. The final manuscript was reviewed by LG. All authors read and approved the final version.



XSL•FO
RenderX

## Conflicts of Interest

None declared.

## References


1. Ndumele CE, Neeland IJ, Tuttle KR, et al. A synopsis of the evidence for the science and clinical management of cardiovascular-kidney-metabolic (CKM) syndrome: a scientific statement from the American Heart Association. Circulation 2023 Nov 14;148(20):1636-1664. [doi: 10.1161/CIR.0000000000001186] [Medline: 37807920]
2. Nakashima A, Kato K, Ohkido I, Yokoo T. Role and treatment of insulin resistance in patients with chronic kidney disease: a review. Nutrients 2021 Dec 2;13(12):4349. [doi: 10.3390/nu13124349] [Medline: 34959901]
3. Lin L, Pan X, Feng Y, Yang J. Chronic kidney disease combined with metabolic syndrome is a non-negligible risk factor. Ther Adv Endocrinol Metab 2024;15:20420188241252309. [doi: 10.1177/20420188241252309] [Medline: 39071115]
4. Pammer LM, Lamina C, Schultheiss UT, et al. Association of the metabolic syndrome with mortality and major adverse cardiac events: a large chronic kidney disease cohort. J Intern Med 2021 Dec;290(6):1219-1232. [doi: 10.1111/joim.13355] [Medline: 34342064]
5. Meikle PJ, Summers SA. Sphingolipids and phospholipids in insulin resistance and related metabolic disorders. Nat Rev Endocrinol 2017 Feb;13(2):79-91. [doi: 10.1038/nrendo.2016.169] [Medline: 27767036]
6. Zewinger S, Schumann T, Fliser D, Speer T. Innate immunity in CKD-associated vascular diseases. Nephrol Dial Transplant 2016 Nov;31(11):1813-1821. [doi: 10.1093/ndt/gfv358] [Medline: 26454223]
7. Kadowaki T, Maegawa H, Watada H, et al. Interconnection between cardiovascular, renal and metabolic disorders: a narrative review with a focus on Japan. Diabetes Obes Metab 2022 Dec;24(12):2283-2296. [doi: 10.1111/dom.14829] [Medline: 35929483]
8. Cersosimo E, DeFronzo RA. Insulin resistance and endothelial dysfunction: the road map to cardiovascular diseases. Diabetes Metab Res Rev 2006;22(6):423-436. [doi: 10.1002/dmrr.634] [Medline: 16506274]
9. Malik S, Wong ND, Franklin SS, et al. Impact of the metabolic syndrome on mortality from coronary heart disease, cardiovascular disease, and all causes in United States adults. Circulation 2004 Sep 7;110(10):1245-1250. [doi: 10.1161/01.CIR.0000140677.20606.0E] [Medline: 15326067]
10. Ostrominski JW, Arnold SV, Butler J, et al. Prevalence and overlap of cardiac, renal, and metabolic conditions in US adults, 1999-2020. JAMA Cardiol 2023 Nov 1;8(11):1050-1060. [doi: 10.1001/jamacardio.2023.3241] [Medline: 37755728]
11. Roth S, M'Pembele R, Matute P, Kotfis K, Larmann J, Lurati Buse G. Cardiovascular-kidney-metabolic syndrome: Association with adverse events after major noncardiac surgery. Anesth Analg 2024 Sep 1;139(3):679-681. [doi: 10.1213/ANE.0000000000006975] [Medline: 39159243]
12. Sarafidis P, Papadopoulos CE, Kamperidis V, Giannakoulas G, Doumas M. Cardiovascular protection with sodium-glucose cotransporter-2 inhibitors and mineralocorticoid receptor antagonists in chronic kidney disease: a milestone achieved. Hypertension 2021 May 5;77(5):1442-1455. [doi: 10.1161/HYPERTENSIONAHA.121.17005] [Medline: 33775130]
13. Major RW, Cheng MRI, Grant RA, et al. Cardiovascular disease risk factors in chronic kidney disease: a systematic review and meta-analysis. PLoS ONE 2018;13(3):e0192895. [doi: 10.1371/journal.pone.0192895] [Medline: 29561894]
14. Feigin VL, Brainin M, Norrving B, et al. World Stroke Organization (WSO): global stroke fact sheet 2022. Int J Stroke 2022 Jan;17(1):18-29. [doi: 10.1177/17474930211065917] [Medline: 34986727]
15. Wu S, Wu B, Liu M, et al. Stroke in China: advances and challenges in epidemiology, prevention, and management. Lancet Neurol 2019 Apr;18(4):394-405. [doi: 10.1016/S1474-4422(18)30500-3] [Medline: 30878104]
16. Charlson FJ, Baxter AJ, Cheng HG, Shidhaye R, Whiteford HA. The burden of mental, neurological, and substance use disorders in China and India: a systematic analysis of community representative epidemiological studies. Lancet 2016 Jul 23;388(10042):376-389. [doi: 10.1016/S0140-6736(16)30590-6] [Medline: 27209143]
17. Flandreau EI, Ressler KJ, Owens MJ, Nemeroff CB. Chronic overexpression of corticotropin-releasing factor from the central amygdala produces HPA axis hyperactivity and behavioral anxiety associated with gene-expression changes in the hippocampus and paraventricular nucleus of the hypothalamus. Psychoneuroendocrinology 2012 Jan;37(1):27-38. [doi: 10.1016/j.psyneuen.2011.04.014] [Medline: 21616602]
18. Li H, Xia N. The role of oxidative stress in cardiovascular disease caused by social isolation and loneliness. Redox Biol 2020 Oct;37:101585. [doi: 10.1016/j.redox.2020.101585] [Medline: 32709420]
19. Taraz M, Taraz S, Dashti-Khavidaki S. Association between depression and inflammatory/anti-inflammatory cytokines in chronic kidney disease and end-stage renal disease patients: a review of literature. Hemodial Int 2015 Jan;19(1):11-22. [doi: 10.1111/hdi.12200] [Medline: 25040322]
20. Hansson GK. Inflammation and atherosclerosis: the end of a controversy. Circulation 2017 Nov 14;136(20):1875-1877. [doi: 10.1161/CIRCULATIONAHA.117.030484] [Medline: 28916641]
21. Barreto DV, Barreto FC, Liabeuf S, et al. Plasma interleukin-6 is independently associated with mortality in both hemodialysis and pre-dialysis patients with chronic kidney disease. Kidney Int 2010 Mar;77(6):550-556. [doi: 10.1038/ki.2009.503] [Medline: 20016471]




XSL•FO
RenderX

22. Gilsanz P, Kubzansky LD, Tchetgen Tchetgen EJ, et al. Changes in depressive symptoms and subsequent risk of stroke in the cardiovascular health study. Stroke 2017 Jan;48(1):43-48. [doi: 10.1161/STROKEAHA.116.013554] [Medline: 27924053]
23. Beekman AT, Deeg DJ, Van Limbeek J, Braam AW, De Vries MZ, Van Tilburg W. Criterion validity of the Center for Epidemiologic Studies Depression scale (CES-D): results from a community-based sample of older subjects in The Netherlands. Psychol Med 1997 Jan;27(1):231-235. [doi: 10.1017/s0033291796003510] [Medline: 9122304]
24. Li W, Shen C, Kong W, et al. Association between the triglyceride glucose-body mass index and future cardiovascular disease risk in a population with cardiovascular-kidney-metabolic syndrome stage 0-3: a nationwide prospective cohort study. Cardiovasc Diabetol 2024 Aug 7;23(1):292. [doi: 10.1186/s12933-024-02352-6] [Medline: 39113004]
25. Ma YC, Zuo L, Chen JH, et al. Modified glomerular filtration rate estimating equation for Chinese patients with chronic kidney disease. J Am Soc Nephrol 2006 Oct;17(10):2937-2944. [doi: 10.1681/ASN.2006040368] [Medline: 16988059]
26. Luo H, Li J, Zhang Q, et al. Obesity and the onset of depressive symptoms among middle-aged and older adults in China: evidence from the CHARLS. BMC Public Health 2018 Jul 24;18(1):909. [doi: 10.1186/s12889-018-5834-6] [Medline: 30041632]
27. Zhou L, Ma X, Wang W. Relationship between cognitive performance and depressive symptoms in Chinese older adults: the China Health and Retirement Longitudinal Study (CHARLS). J Affect Disord 2021 Feb 15;281:454-458. [doi: 10.1016/j.jad.2020.12.059] [Medline: 33360747]
28. Ersoy B, Herzog ML, Pan W, et al. The atypical antidepressant tianeptine confers neuroprotection against oxygen-glucose deprivation. Eur Arch Psychiatry Clin Neurosci 2024 Jun;274(4):777-791. [doi: 10.1007/s00406-023-01685-9] [Medline: 37653354]
29. Kim S, Park ES, Chen PR, Kim E. Dysregulated hypothalamic–pituitary–adrenal axis is associated with increased inflammation and worse outcomes after ischemic stroke in diabetic mice. Front Immunol 2022;13:864858. [doi: 10.3389/fimmu.2022.864858]
30. Hassamal S. Chronic stress, neuroinflammation, and depression: an overview of pathophysiological mechanisms and emerging anti-inflammatories. Front Psychiatry 2023;14:1130989. [doi: 10.3389/fpsyt.2023.1130989] [Medline: 37252156]
31. Koutsaliaris IK, Moschonas IC, Pechlivani LM, Tsouka AN, Tselepis AD. Inflammation, oxidative stress, vascular aging and atherosclerotic ischemic stroke. Curr Med Chem 2022;29(34):5496-5509. [doi: 10.2174/0929867328666210921161711] [Medline: 34547993]
32. Höglund E, Øverli Ø, Winberg S. Tryptophan metabolic pathways and brain serotonergic activity: a comparative review. Front Endocrinol (Lausanne) 2019;10:158. [doi: 10.3389/fendo.2019.00158] [Medline: 31024440]
33. Ko M, Choi-Kwon S, Jun SE, et al. Poststroke emotional disturbances and a tryptophan hydroxylase 2 gene polymorphism. Brain Behav 2018 Feb;8(2):e00892. [doi: 10.1002/brb3.892] [Medline: 29484259]
34. Jayakumar S, Jennings S, Halvorsrud K, et al. A systematic review and meta-analysis of the evidence on inflammation in depressive illness and symptoms in chronic and end-stage kidney disease. Psychol Med 2023 Sep;53(12):5839-5851. [doi: 10.1017/S0033291722003099] [Medline: 36254747]
35. Li X, Xiang X, Hu J, et al. Association between serum cortisol and chronic kidney disease in patients with essential hypertension. Kidney Blood Press Res 2016;41(4):384-391. [doi: 10.1159/000443435] [Medline: 27344357]
36. Sung CC, Hsu YC, Chen CC, Lin YF, Wu CC. Oxidative stress and nucleic acid oxidation in patients with chronic kidney disease. Oxid Med Cell Longev 2013;2013:301982. [doi: 10.1155/2013/301982] [Medline: 24058721]
37. Spoto B, Di Betta E, Mattace-Raso F, et al. Pro- and anti-inflammatory cytokine gene expression in subcutaneous and visceral fat in severe obesity. Nutr Metab Cardiovasc Dis 2014 Oct;24(10):1137-1143. [doi: 10.1016/j.numecd.2014.04.017] [Medline: 24984824]
38. Chao X, Fang Y, Lu Z, et al. Impairments of neurovascular coupling after stroke lower glymphatic system function and lead to depressive symptom: a longitudinal cohort study. J Affect Disord 2024 Dec 15;367:255-262. [doi: 10.1016/j.jad.2024.08.229] [Medline: 39236880]
39. Liu L, Xu M, Marshall IJ, Wolfe CD, Wang Y, O'Connell MD. Prevalence and natural history of depression after stroke: a systematic review and meta-analysis of observational studies. PLoS Med 2023 Mar;20(3):e1004200. [doi: 10.1371/journal.pmed.1004200] [Medline: 36976794]
40. Gilsanz P, Walter S, Tchetgen Tchetgen EJ, et al. Changes in depressive symptoms and incidence of first stroke among middle‐aged and older US adults. J Am Heart Assoc 2015 May 13;4(5):e001923. [doi: 10.1161/JAHA.115.001923] [Medline: 25971438]
41. Pan A, Sun Q, Okereke OI, Rexrode KM, Hu FB. Depression and risk of stroke morbidity and mortality: a meta-analysis and systematic review. JAMA 2011 Sep 21;306(11):1241-1249. [doi: 10.1001/jama.2011.1282] [Medline: 21934057]
42. Cui Y, Zhu C, Lian Z, et al. Prospective association between depressive symptoms and stroke risk among middle-aged and older Chinese. BMC Psychiatry 2021 Oct 27;21(1):532. [doi: 10.1186/s12888-021-03492-9] [Medline: 34706692]
43. Chen Y, Bennett D, Clarke R, et al. Patterns and correlates of major depression in Chinese adults: a cross-sectional study of 0.5 million men and women. Psychol Med 2017 Apr;47(5):958-970. [doi: 10.1017/S0033291716002889] [Medline: 27919307]




XSL•FO
RenderX

44. Phillips MR, Zhang J, Shi Q, et al. Prevalence, treatment, and associated disability of mental disorders in four provinces in China during 2001–05: an epidemiological survey. The Lancet 2009 Jun;373(9680):2041-2053. [doi: 10.1016/S0140-6736(09)60660-7]

## Abbreviations

**AHA:** American Heart Association
**CHARLS:** China Health and Retirement Longitudinal Study
**CKD:** chronic kidney disease
**CKM:** cardiovascular-kidney-metabolic
**CRP:** C-reactive protein
**CVD:** cardiovascular disease
**eGFR:** estimated glomerular filtration rate
**$HbA_{1c}$:** hemoglobin $A_{1c}$
**HDL:** high-density lipoprotein
**HR:** hazard ratio
**LDL:** low-density lipoprotein
**RCS:** restricted cubic spline
**TC:** total cholesterol
**UA:** uric acid

*Edited by M Balcarras, N Cahill; submitted 20.Jul.2025; peer-reviewed by ADERONKEELIZABETH FAKAYODE, Y Chen, Y Zhang; accepted 22.Jan.2026; published 01.Apr.2026.*

*Please cite as:*
*Li Y, Wang L, Gao L*
*Association Between Depressive Symptoms and Incidence of Stroke in a Population With Cardiovascular-Kidney-Metabolic Syndrome Stages 0 to 3: Nationwide Prospective Cohort Study*
*Interact J Med Res 2026;15:e80988*
*URL: https://www.i-jmr.org/2026/1/e80988*
*doi:10.2196/80988*

© Yunjie Li, Lin Wang, Lujing Gao. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 1.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Patient-Reported Quality of Life in Lung Transplantation: Prospective Cohort Study

Peng Li[1], PhD; Jiawei Wang[1], MD; Hongyi Wang[1], MD; Shuo Li[1], MD; Lei Wang[1], MD; Shan Gao[1], MD; Chunxiao Hu[2], MD; Guangjian Zhang[1], MD, PhD

[1]Department of Thoracic Surgery, First Affiliated Hospital of Xi'an Jiaotong University, Yanta West Road, Xi'an, China

[2]Department of Anesthesiology, The Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi, China

**Corresponding Author:**
Guangjian Zhang, MD, PhD
Department of Thoracic Surgery, First Affiliated Hospital of Xi'an Jiaotong University, Yanta West Road, Xi'an, China

## Abstract

**Background:** Lung transplantation (LTx) is an established treatment for patients with end-stage lung diseases and can substantially improve survival. However, posttransplant recovery involves complex physical, psychological, and functional challenges, making quality of life (QoL) an important outcome beyond survival alone. Evidence on early longitudinal QoL changes after LTx remains limited, particularly in Chinese recipients and when assessed using disease-specific patient-reported outcome instruments.

**Objective:** This study aimed to assess QoL and track its longitudinal changes in patients who underwent LTx using a disease-specific patient-reported outcome instrument.

**Methods:** This single-center prospective cohort study screened 66 patients for enrollment who underwent LTx at the Xi'an Jiaotong University Lung Transplantation Center in China from June 2023 to December 2025. QoL was assessed using the Chinese version of the Lung Transplant Quality of Life questionnaire at baseline (preoperative) and serially at 1, 2, 3, 4, 5, and 6 months after surgery. The Lung Transplant Quality of Life questionnaire comprises 40 items across 7 domains, and each domain score was calculated as the item mean (range 0 - 4). Random-intercept linear mixed models (LMMs) with both categorical and linear time specifications were fitted separately for each domain to evaluate longitudinal changes, accounting for repeated measures and incomplete follow-up under the missing at random assumption.

**Results:** A total of 203 interviews from 51 eligible patients were included in this study, with a median of 5 (IQR 2-6) interviews for each patient. LMM analysis revealed significant overall time effects in 6 of 7 domains (likelihood ratio test: $P<.01$ in all cases). Health perceptions showed the earliest and largest improvement (LMM-adjusted mean difference [MD] −1.58 at month 6; β=−0.23 points per month; $P<.001$), followed by respiratory symptoms (MD −0.95 at month 6; β=−0.15 per month), anxiety and depression (MD −0.82 at month 6; β=−0.13 per month), digestive symptoms (MD −0.50 at month 6; β=−0.10 per month), and cognitive limitations (MD −0.51 at month 6; β=−0.09 per month; $P<.001$ in all cases). Global well-being was the only domain with a positive trajectory (MD 1.18 at month 6; β=0.13 per month; $P=.002$), with significant improvement already evident at month 1 (MD 0.71; $P=.23$).

**Conclusions:** LTx significantly improved QoL in patients with end-stage lung diseases, with the most substantial gains in health perceptions and respiratory symptoms. Patient-reported outcome–based monitoring can capture clinically meaningful QoL changes and should be integrated into routine posttransplant care to identify critical intervention time points and optimize long-term recovery.

***(Interact J Med Res 2026;15:e86211)*** doi:10.2196/86211

**KEYWORDS**

lung transplantation; quality of life; patient-reported outcomes; PROs; symptoms

## Introduction

Lung transplantation (LTx) is the definitive treatment for end-stage lung diseases, including chronic obstructive lung disease, idiopathic pulmonary fibrosis, and cystic fibrosis [1,2]. With advancements in surgical techniques and perioperative care, the 1- and 5-year survival rates have exceeded 80% and 70%, respectively, although outcomes vary across transplantation centers [1]. Despite these improvements in survival, patients after LTx continue to face multiple physical and psychological complications at various time points, including primary graft dysfunction [3], acute kidney injury [4], atrial arrhythmias [5], anastomotic fistulas [6], fungal infections [7], airway hemorrhage [8], and delirium, all of which severely undermine postoperative survival and quality of life (QoL) [9]. Current clinical practices predominantly focus on reducing



XSL•FO
RenderX

postoperative complications, improving survival rates, and prolonging survival time [10,11], which may overlook the patient perspective on life quality.

This study was motivated by concerns raised by patients and their families, who frequently asked whether their symptoms and functional impairments would resolve and how these would change over time after surviving surgery. A limited body of literature has documented psychosocial and mental health outcomes after transplantation, but these studies have primarily focused on liver, heart, or kidney transplantation [12-14]. Furthermore, the assessment of QoL in patients undergoing LTx has largely relied on generic scales such as the 36-Item Short Form Health Survey or EQ-5D [15,16], which may not adequately capture the specific symptoms, functions, and lived experiences following LTx. Singer et al [17] developed the Lung Transplant Quality of Life (LT-QoL) scale in 2019, a disease-specific instrument that was subsequently translated and validated for the Chinese cultural context, thereby facilitating more accurate QoL assessment in this population.

Patient-reported outcomes (PROs) have been increasingly applied in clinical practice to monitor patient symptoms and experiences [18], improve clinical care [19], and support drug development evaluations by regulatory agencies [20]. However, very little is currently known about longitudinal QoL changes and trajectories in patients undergoing LTx based on PROs. Therefore, we conducted a prospective cohort study to assess QoL changes in patients undergoing LTx at our center using the Chinese version of the LT-QoL questionnaire and characterize the temporal patterns of recovery across multiple QoL domains.

## Methods

### Study Design and Patients

This was a single-center prospective cohort study conducted at the Xi'an Jiaotong University Lung Transplantation Center (First Affiliated Hospital of Xi'an Jiaotong University, China). A total of 66 patients who underwent LTx from June 2023 to December 2025 were screened for enrollment. Of these 66 patients, 58 (87.9%) patients or their families were fully informed about the study procedures, whereas 8 (12.1%) declined participation. Of the 58 initial patients, 7 (12.1%) were subsequently excluded (n=1 due to retransplantation and n=6 who withdrew due to being too weak to self-report), producing a final analytic cohort of 51 (87.9%) patients who contributed 203 valid interviews across 7 time points.

Inclusion criteria were (1) age of 18 years or above, (2) having undergone LTx, and (3) ability to understand the study content. Patients were excluded if they (1) underwent retransplantation or multi-organ transplantation, (2) were too weak to self-report outcomes, or (3) were diagnosed with active psychosis or severe cognitive impairment during follow-up.

### Ethical Considerations

This study was approved by the institutional review board of the First Affiliated Hospital of Xi'an Jiaotong University, China (approval number: LLSBPJ-2023-474), was exempted from clinical trial registration as it was an observational study, and was conducted in accordance with the Declaration of Helsinki. A total of 49.0% (25/51) of the patients provided written informed consent before surgery, whereas the remaining patients provided verbal informed consent due to the urgency of surgery and subsequently supplemented it with written consent at the earliest appropriate time.

### Procedures

All enrolled patients were invited to complete the LT-QoL scale at each scheduled assessment. The LT-QoL is a disease-specific questionnaire developed for the assessment of QoL in patients undergoing LTx [17]. A Chinese version was authorized and demonstrated adequate validity, reliability, and internal consistency. It consists of 40 items assessing 7 domains grouped under 4 overarching categories: symptoms (respiratory symptoms: 9 items; digestive symptoms: 8 items), health perceptions (5 items), functional impairment (sexual problems: 3 items; cognitive limitations: 4 items), and well-being (anxiety and depression: 9 items; global well-being: 2 items). Each item is quantified on a 5-point Likert scale (0 - 4), with higher scores indicating greater symptom burden or impairment for all domains except global well-being, where higher scores represent better outcomes. For each domain, a composite score was calculated as the arithmetic mean of its constituent items (item mean score; range 0 - 4). Further details are provided in Multimedia Appendix 1.

For each patient, assessments were scheduled before surgery (T0; baseline) and at 1, 2, 3, 4, 5, and 6 months after LTx (T1-T6). For patients who were unable to self-report, family members or caregivers completed the questionnaire based on the patient's actual condition. Assessments were conducted in the LTx specialized outpatient clinic or via WeChat by a trained study team comprising 2 specialized nurses, 2 researchers, and 2 clinicians. Each visit could take place within a window of –7 days to +7 days from the actual scheduled appointment. Data collection was performed by a nurse under the supervision of a researcher and a clinician to ensure accuracy and completeness.

### Data Collection and Management

In this longitudinal study, data were collected and managed using the REDCap (Research Electronic Data Capture; Vanderbilt University) platform hosted at Xi'an Jiaotong University Health Science Center [21]. REDCap is a secure, web-based application designed to support data capture for research studies [22]. Patients could choose to complete either an electronic or paper-based questionnaire. Paper-based questionnaire data were entered into REDCap and verified by an independent nurse.

Demographic information and clinical characteristics were obtained from medical records supplemented by self-reported information about demographics, health behaviors, and medical history. All data were double-checked and cleaned by a professional statistician and a clinician to ensure accuracy and maintain high data quality.

### Statistical Analysis

Descriptive statistics, including means with SDs and medians with IQRs, were used to summarize patient characteristics.



XSL•FO
RenderX

Continuous variables were compared using unpaired 2-tailed Student *t* tests or Wilcoxon signed-rank tests for normally and nonnormally distributed data, respectively. Categorical variables were presented as frequencies and percentages and compared using Fisher exact tests or chi-square tests as appropriate. A 2-sided *P* value of less than .05 was considered statistically significant.

To evaluate longitudinal changes in QoL scores, we used random-intercept linear mixed models (LMMs) fitted separately for each of the 7 QoL domains using restricted maximum likelihood estimation. Two complementary model specifications were used for each domain. In model 1 (categorical time), score was modeled as a function of time point + (1 | participant), where time point was treated as a categorical fixed effect with T0 (baseline) as the reference level. This model yielded the LMM-estimated marginal mean at each time point and the adjusted mean difference (MD) vs baseline with 95% CIs and Satterthwaite-approximated *P* values for each pairwise contrast. In model 2 (linear time), score was modeled as a function of month + (1 | participant), where month was treated as a continuous fixed effect (0, 1, 2,..., 6). This model estimated the average linear slope (β) per month with its 95% CI, quantifying the overall rate of change across the 6-month follow-up period.

The random-intercept specification accounted for within-patient correlation due to repeated measures and accommodated the unbalanced longitudinal design (51 patients contributing 203 interviews). All available observations were used under the missing at random (MAR) assumption, thereby mitigating potential bias from incomplete follow-up data. Individual patient trajectories overlaid with the LMM population fit are shown in Figure S1 in Multimedia Appendix 1, illustrating the heterogeneity among patients and the pattern of missing data that motivated the mixed-effects approach. The intraclass correlation coefficient (ICC; $\sigma^2_u/[\sigma u^2_u + \sigma^2_e]$) was calculated for each domain to quantify the proportion of the total variance attributable to between-patient differences.

The overall significance of the time effect for each domain was assessed using a likelihood ratio test (LRT) comparing the categorical time model against a null model with random intercept only (score modeled as a function of 1 + [1 | participant]). Estimated marginal means with 95% CIs were derived from model 1 for each domain at each time point. Adjusted MDs vs baseline with 95% Wald CIs were displayed as a forest plot; the linear slopes per month from model 2 were appended to the same forest plot for each domain. Individual patient trajectories were overlaid with the LMM population fit in a spaghetti plot (Figure S1 in Multimedia Appendix 1) to illustrate individual-level heterogeneity and the pattern of missing data. In addition to the domain-level analyses, item-level LMMs (score modeled as a function of time point + [1 | participant]) were fitted for all 40 individual LT-QoL items to identify specific symptoms or concerns driving the observed domain-level changes.

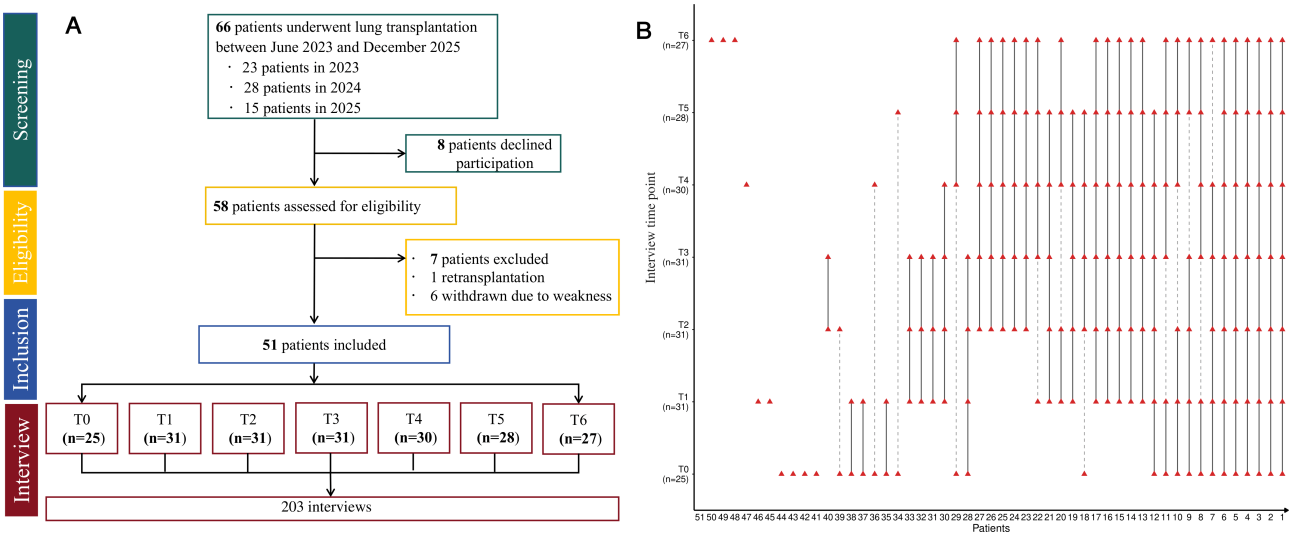
All analyses were performed using R (version 4.4.2; R Foundation for Statistical Computing) with the *lme4* package for LMM estimation, the *lmerTest* package for Satterthwaite df and *P* values, and the *emmeans* package for estimated marginal means and pairwise contrasts.

# Results

## Clinical Characteristics

Of the 66 patients who underwent LTx at our center, 15 (22.7%) were excluded: 8 (53.3%) of them declined participation, 1 (6.7%) underwent retransplantation, and 6 (40%) withdrew consent. A total of 51 eligible patients completed 203 valid QoL interviews from baseline to 6 months postoperatively. The median number of interviews per patient was 5 (IQR 2 - 6), and 80.4% (41/51) and 66.7% (34/51) of the patients completed at least 2 and 3 valid interviews, respectively (Figure 1).

**Figure 1.** Patient enrollment and interview completion: (A) flowchart of patient screening, eligibility assessment, and inclusion and (B) interview completion for each patient across time points. Solid lines indicate consecutive follow-up visits; dashed lines indicate gaps due to missing interviews. T0: baseline (before surgery); T1 to T6: first to sixth month after lung transplantation.






XSL•FO
RenderX

The mean age of the eligible patients was 56.4 (SD 9.3) years, and most were male (44/51, 86.3%) and did not have a history of smoking (39/51, 76.5%). Few of them self-reported current smoking, alcohol use, or major comorbidities (hypertension: 5/51, 9.8%; diabetes: 2/51, 3.9%; cerebrovascular or cardiovascular disease: 4/51, 7.8%). The most frequent pretransplant diagnosis was idiopathic pulmonary fibrosis (26/51, 51.0%), followed by chronic obstructive pulmonary disease (13/51, 25.5%) and pneumoconiosis (10/51, 19.6%). Mean BMI was 20.5 (SD 3.7) kg/m$^2$, indicating a generally lean cohort consistent with end-stage lung disease. Bilateral sequential LTx was the most common surgery type (27/51, 52.9%), followed by single transplantation (Table 1).

**Table .** Demographic and clinical characteristics of the patients (N=51).

| Characteristic | Values |
|---|---|
| Age (y), mean (SD) | 56.4 (9.3) |
| Sex, n (%) | |
| Female | 7 (13.7) |
| Male | 44 (86.3) |
| BMI (kg/m$^2$), mean (SD) | 20.5 (3.7) |
| Diabetes, n (%) | 2 (3.9) |
| CCD[a], n (%) | 4 (7.8) |
| Smoking history, n (%) | 12 (23.5) |
| Hypertension, n (%) | 5 (9.8) |
| Diagnosis, n (%) | |
| COPD[b] | 13 (25.5) |
| IPF[c] | 26 (51.0) |
| Pneumoconiosis | 10 (19.6) |
| Others | 2 (3.9) |
| Surgery type, n (%) | |
| Single LTx[d] | 24 (47.1) |
| Bilateral LTx | 27 (52.9) |
| ECMO[e] type, n (%) | |
| VV[f] | 27 (52.9) |
| VA[g] | 12 (23.5) |
| VAV[h] | 2 (3.9) |

[a]CCD: cardiovascular or cerebrovascular disease.
[b]COPD: chronic obstructive pulmonary disease.
[c]IPF: idiopathic pulmonary fibrosis.
[d]LTx: lung transplantation.
[e]ECMO: extracorporeal membrane oxygenation.
[f]VV: veno-venous ECMO.
[g]VA: veno-arterial ECMO.
[h]VAV: veno-arterial-veno ECMO.

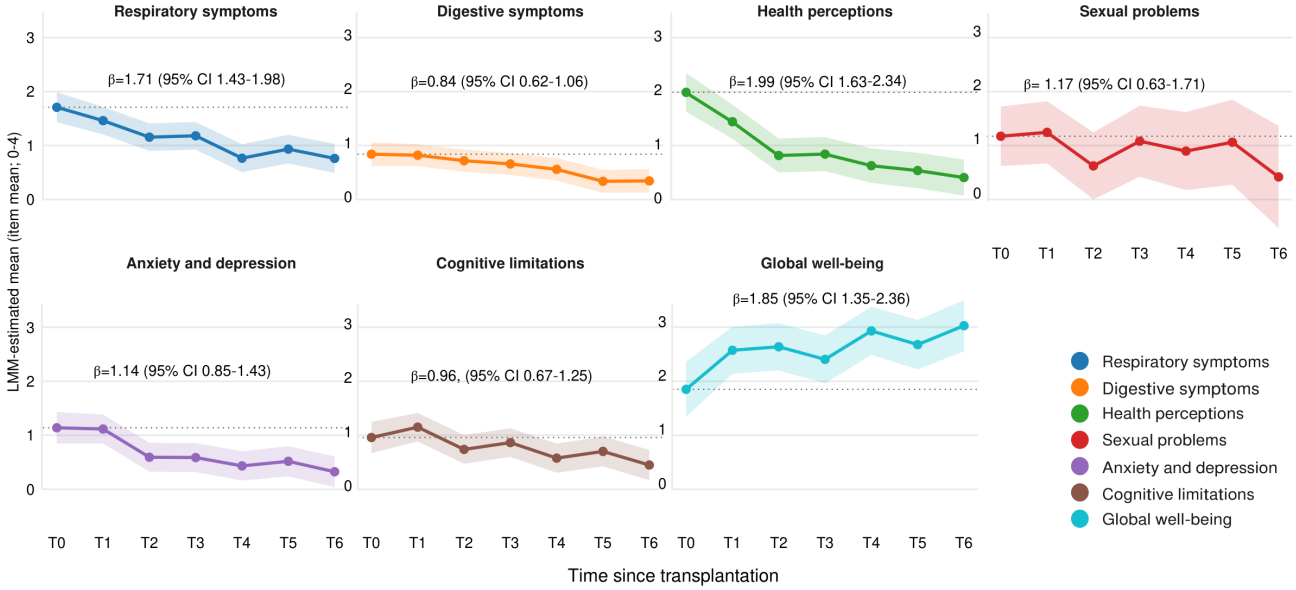
## Respiratory Symptoms

In the random-intercept LMM, the model-estimated baseline mean for the respiratory symptom domain was 1.71 points (SD 0.14; 95% CI 1.43-1.98) on the item mean scale from 0 to 4 (Figure 2). The overall time effect was significant (LRT: $\chi^2_6$=47.8; $P$<.001), and the ICC was 0.40, confirming that repeated measures within patients were substantially correlated and that the mixed-effects approach was appropriate (Table S1 in Multimedia Appendix 1).



XSL•FO
RenderX

**Figure 2.** Linear mixed model (LMM)–estimated marginal means of quality of life domains over 6 months after lung transplantation. Random-intercept LMMs were fitted separately for each domain. Shaded bands represent 95% Wald CIs. Lower scores indicate better status in all domains except global well-being, for which higher scores represent better outcomes. T0: baseline (before surgery); T1 to T6: first to sixth month after lung transplantation.

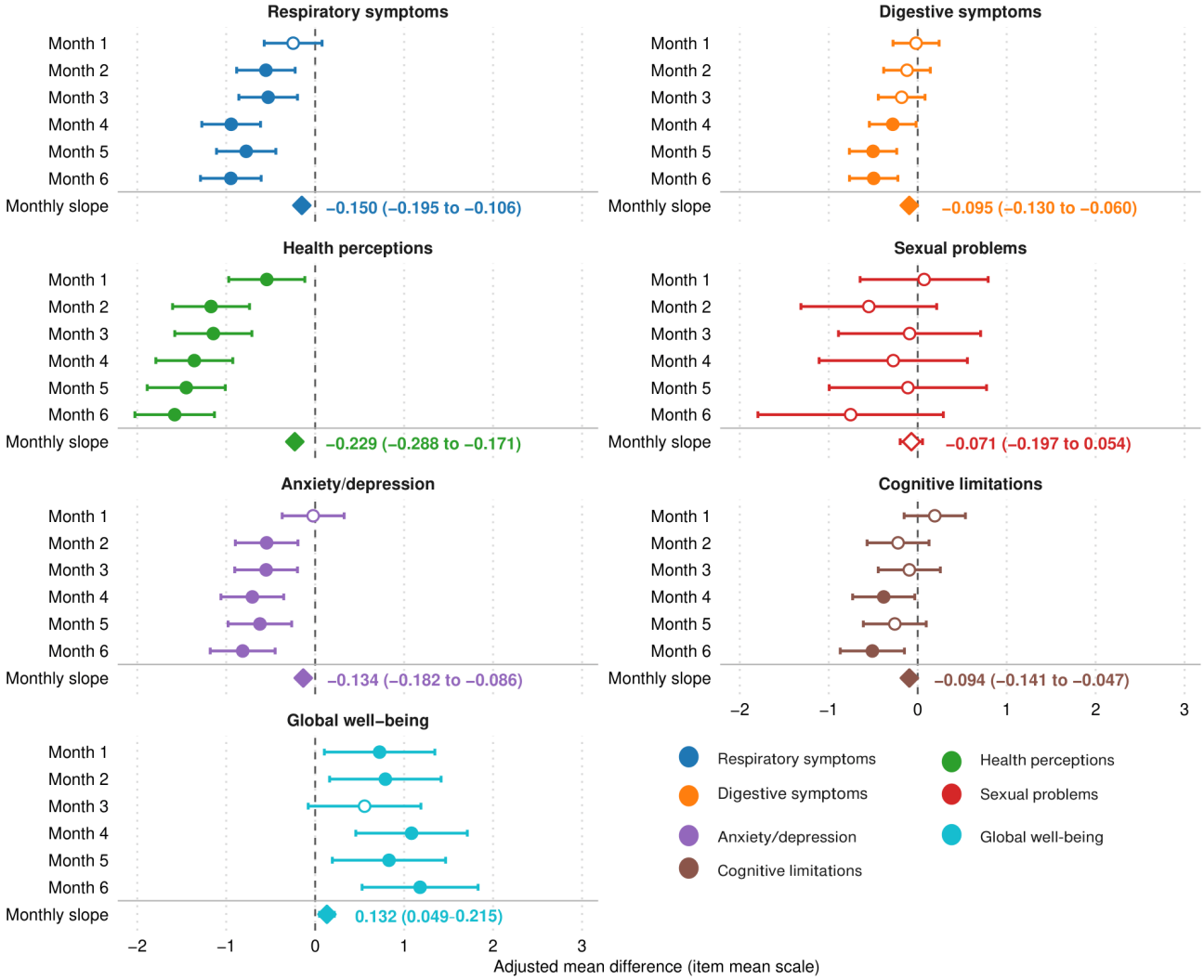



The adjusted MD vs baseline was not significant at month 1 (MD −0.25, 95% CI −0.57 to 0.07; *P*=.13) but became significant from month 2 onward (MD −0.57, 95% CI −0.89 to −0.24; *P*=.001). The improvement continued to increase, reaching an MD of −0.95 (95% CI −1.29 to −0.61; *P*<.001) at month 6. The linear slope was −0.15 points per month (95% CI −0.20 to −0.11; *P*<.001), indicating a steady and clinically meaningful decline in respiratory symptoms over the 6-month follow-up period (Figure 3).



XSL•FO
RenderX

**Figure 3.** Forest plot of linear mixed model–adjusted mean differences vs baseline and linear slopes per month by quality of life domain. Circles denote time point contrasts vs baseline (model 1); diamonds denote the linear slope per month (model 2). Filled markers indicate a *P* value of .05 or less; blank markers indicate a *P* value above .05. Error bars represent 95% Wald CIs. The dashed vertical line marks the null value of 0 (no change).




At the item level (Table S2 in Multimedia Appendix 1), the most burdensome preoperative respiratory symptoms were cough with sputum (baseline $\beta_0$=2.79), sputum production (baseline $\beta_0$=2.75), wheezing (baseline $\beta_0$=2.71), and shortness of breath (baseline $\beta_0$=2.65). These items also showed the largest LMM-adjusted 6-month improvements (shortness of breath: MD −1.65 and *P*<.001; wheezing: MD −1.58 and *P*<.001; sputum production: MD −1.22 and *P*<.001; cough: MD −1.18 and *P*<.001). Chest tightness (baseline $\beta_0$=2.05) showed no early change (month 1: MD −0.40; *P*=.24) but a significant improvement at month 6 (MD −1.45; *P*<.001). Notably, chest pain showed a transient worsening at month 1 (MD 0.70; *P*=.02) before declining, consistent with the typical early postoperative course; however, this improvement did not reach statistical significance by month 6 (MD −0.52; *P*=.07).

## Digestive Symptoms

The estimated baseline mean for the digestive symptom domain was 0.84 points (SD 0.11; 95% CI 0.62-1.06), substantially lower than that of the respiratory symptom domain. The overall time effect was significant (LRT: $\chi^2_6$=28.0; *P*<.001; ICC=0.41; Table S1 in Multimedia Appendix 1).

Digestive symptoms followed a slower improvement trajectory than respiratory symptoms. No significant change was detected at months 1 to 3 (*P*>.15 in all cases; Figure 2). The adjusted MD first became significant at month 4 (MD −0.28, 95% CI −0.54 to −0.02; *P*=.03) and strengthened throughout months 5 and 6 (month 5: MD −0.50, 95% CI −0.77 to −0.24, and *P*<.001; month 6: MD −0.50, 95% CI −0.77 to −0.22, and *P*<.001). The slope was −0.10 points per month (95% CI −0.13 to −0.06; *P*<.001).

At the item level (Table S2 in Multimedia Appendix 1), the most burdensome baseline digestive symptom was loss of appetite (baseline $\beta_0$=1.66), which also showed the largest 6-month improvement (MD −1.09; *P*=.001). Fear of being far from a toilet showed both an early improvement at month 1 (MD −0.58; *P*=.005) and a sustained 6-month reduction (MD −0.77; *P*<.001). Altered taste improved significantly by month 6 (MD −0.75; *P*=.02). In contrast, stomach cramps and pain showed a transient worsening at month 1 (MD 0.45; *P*=.02) before returning to baseline levels.



XSL•FO
RenderX

### Cognitive Limitations

The estimated baseline mean for the cognitive limitation domain was 0.96 points (SD 0.15; 95% CI 0.67-1.25), with a significant overall time effect (LRT: $\chi^2_6$=22.4; *P*=.001; ICC=0.36; Table S1 in Multimedia Appendix 1). As shown in Figures 2 and 3, the trajectory of improvement was not significant at months 1 to 3 (month 1: MD 0.19 and *P*=.17; month 3: MD −0.11 and *P*=.18). A significant reduction was first observed at month 4 (MD −0.39, 95% CI −0.73 to −0.04; *P*=.003) and persisted at month 6 (MD −0.51, 95% CI −0.87 to −0.15; *P*=.006). The slope was −0.09 points per month (95% CI −0.14 to −0.05; *P*<.001). Item-level analyses showed the largest 6-month improvements in difficulty sustaining attention (MD −0.80; *P*<.001; Table S2 in Multimedia Appendix 1) and difficulty concentrating and thinking (MD −0.50; *P*=.01), followed by slowed reactions to past events (MD −0.57; *P*=.02). General forgetfulness showed only a nonsignificant numerical decline (MD −0.25; *P*=.35).

### Sexual Problems

For the sexual problem domain, the LMM-estimated baseline mean was 1.17 points (SD 0.28; 95% CI 0.63-1.71). The overall time effect was not significant (LRT: $\chi^2_6$=5.3; *P*=.55; Table S1 in Multimedia Appendix 1), and no individual postoperative time point reached statistical significance vs baseline (*P*>.15 in all cases; Figures 2 and 3). The slope was −0.07 points per month (95% CI −0.20 to 0.05; *P*=.27). The wide CIs reflected substantial missingness in this domain. At the item level, none of the 3 sexual problem items showed significant changes at any time point (*P*>.08 in all cases).

### Health Perceptions

The health perception domain showed the largest and earliest improvement of all 7 domains. The LMM-estimated baseline mean was 1.99 points (SD 0.18; 95% CI 1.63-2.34), the highest baseline score of any domain, indicating that concerns about health status were the most prominent source of preoperative distress. The overall time effect was highly significant (LRT: $\chi^2_6$=65.1; *P*<.001; ICC=0.34; Table S1 in Multimedia Appendix 1).

Crucially, a significant reduction was already clear at month 1 (MD −0.55, 95% CI −0.97 to −0.12; *P*=.01), and the improvement increased monotonically thereafter, reaching an MD of −1.58 (95% CI −2.02 to −1.14; *P*<.001) at month 6 (Figures 2 and 3). The slope was the steepest among all 7 domains: −0.23 points per month (95% CI −0.29 to −0.17; *P*<.001). At the item level (Table S2 in Multimedia Appendix 1), all 5 health perception items improved significantly by month 6, such as fear of infection (MD −1.77; *P*<.001), uncertainty about future health (MD −1.76; *P*<.001), worry about worsening health (MD −1.73; *P*<.001), worry about graft failure or rejection (MD −1.68; *P*<.001), and difficulty planning for the future (MD −0.97; *P*<.001). Notably, uncertainty about future health and worry about graft rejection were already significantly lower at month 1 (MD −0.99 with *P*<.001 and MD −0.55 with *P*=.03, respectively), whereas fear of infection improved more gradually and only reached significance in month 6 (MD −1.76; *P*<.001).

### Anxiety and Depression

The estimated baseline mean for the anxiety and depression domain was 1.14 points (SD 0.15; 95% CI 0.85-1.43). The overall time effect was significant (LRT: $\chi^2_6$=37.1; *P*<.001; ICC=0.36; Table S1 in Multimedia Appendix 1). As with respiratory symptoms, there was no detectable change at month 1 (MD −0.02; *P*=.89), but a significant improvement emerged from month 2 onward (MD −0.56, 95% CI −0.90 to −0.21; *P*=.003) and continued up to month 6 (MD −0.82, 95% CI −1.18 to −0.46; *P*<.001). The slope was −0.13 points per month (95% CI −0.18 to −0.09; *P*<.001).

Item-level analyses showed that the largest 6-month reductions were in uncontrollable worry (MD −1.27; *P*<.001), low mood (MD −1.09; *P*<.001), feeling sad (MD −0.99; *P*<.001), and irritability (MD −0.97; *P*<.001). The items reflecting more severe psychological distress, including inability to stop worrying (MD −0.64; *P*=.003), feeling guilty (MD −0.66; *P*=.001), and feeling like a burden to others (MD −0.66; *P*=.001), also improved significantly, suggesting a broad amelioration of emotional distress (Table S2 in Multimedia Appendix 1). Difficulty sleeping improved significantly by month 6 (MD −0.67; *P*=.02), although there was a nonsignificant trend toward worsening at month 1 (MD 0.51; *P*=.06). Perceived loss of interest in daily activities did not reach significance at month 6 (MD −0.41; *P*=.10).

### Global Well-Being

The global well-being domain (higher scores indicate better outcomes) was the only domain with a positive trajectory. The estimated baseline mean was 1.85 points (SD 0.26; 95% CI 1.35-2.36), and the overall time effect was significant (LRT: $\chi^2_6$=15.7; *P*=.01; ICC=0.27; Table S1 in Multimedia Appendix 1).

The adjusted MD rose to 0.71 (95% CI 0.09-1.33; *P*=.23) at month 1, indicating that patients already perceived meaningful improvement in life quality within the first month after transplantation. The improvement peaked at month 4 (MD 1.01, 95% CI 0.46-1.71; *P*<.001) and was sustained up to month 6 (MD 1.18, 95% CI 0.53-1.83; *P*<.001). The slope was 0.13 points per month (95% CI 0.05-0.22; *P*=.002). At the item level, enjoyment of life showed a borderline improvement at month 1 (MD 0.60; *P*=.08) that became highly significant by month 6 (MD 1.27; *P*<.001). Satisfaction with current QoL improved earlier and more robustly: it was already significantly higher at month 1 (MD 0.89; *P*=.02) and continued to improve up to month 6 (MD 1.11; *P*=.004), suggesting that patients' subjective life satisfaction recovered more rapidly than their ability to enjoy daily activities following LTx.

## Discussion

### Principal Findings

In this study, we found significant improvements in the QoL of patients undergoing LTx in the perioperative period, with the



XSL•FO
RenderX

earliest and largest improvement in health perceptions followed by respiratory symptoms and anxiety and depression. In contrast, the sexual problem domain showed no significant change over time, partly attributable to high missingness and cultural reporting barriers. These findings demonstrate that LTx meaningfully improves overall QoL for patients with end-stage lung diseases across multiple domains, with clinically important reductions in symptom burden, psychological distress, and health-related concerns.

Respiratory symptoms were the predominant preoperative concern, with an estimated baseline mean of 1.71 points (95% CI 1.43-1.98), and the overall time effect was highly significant. This finding is consistent with those of previous studies linking forced vital capacity and forced expiratory volume improvements with pulmonary function recovery after transplantation [23,24]. Unlike previous studies that primarily focused on pulmonary function [25,26], we also monitored gastrointestinal symptom changes using LMMs. The digestive symptom domain showed a slower improvement trajectory than respiratory symptoms: no significant change was detected until month 4, with the improvement strengthening up to month 6. A possible explanation for this improvement may be the resolution of hypoxia and the subsequent enhancement in overall health, which collectively contribute to the restoration of normal metabolic and digestive functions [27]. Notably, stomach cramps and pain showed a transient worsening at month 1 before returning to baseline levels, and chest pain similarly worsened early. This early symptom exacerbation pattern likely reflects the combined effects of surgical stress [28], immunosuppressive medications (eg, steroids and calcineurin inhibitors) [29], and associated gastrointestinal complications. Similar gastrointestinal symptom profiles have been reported in kidney transplant recipients, with strong associations with poorer QoL [30]. Our findings support the integration of routine gastrointestinal symptom monitoring and proactive management into posttransplant care protocols.

Postoperative cognitive impairment is a common clinical phenomenon with a reported incidence of 58.3%, influenced by factors such as operation time, intensive care unit stay, pain, and cold ischemia time [31,32]. Decreases in processing speed and executive function at 6 months may signify worse survival and higher occurrence of chronic lung allograft dysfunction [33]. In contrast, the sexual problem domain was the only domain for which the LMM detected no significant overall time effect, and no individual postoperative time point reached significance vs baseline. In traditional Chinese culture, sexual health remains a sensitive and often stigmatized topic, which may lead to underreporting or social desirability bias. We did not apply analytical adjustments for potential underreporting as the degree cannot be quantified from the available data. Therefore, this null finding should be interpreted with caution. Compared to liver or kidney transplantation, where sexual problems are more prevalent [34,35], the sexual health domain in LTx remains understudied, and future research should consider culturally sensitive approaches such as anonymous electronic surveys, gender-matched interviewers, or validated proxy measures to improve assessment accuracy in Chinese transplant populations.

The health perception domain showed the earliest and largest improvement of all 7 domains, indicating that health-related concerns were the most prominent source of preoperative distress. The gradual resolution of these concerns suggests that patients experienced progressive confidence recovery. However, the persistent concern about infections despite decreasing scores underscores the challenges of living with immunosuppressive therapy. More than 30% of patients undergoing liver transplant have been reported to experience clinically significant mental distress after surgery [36], and our findings reinforce the importance of sustained psychological support throughout recovery. For the anxiety and depression domain, the adjusted improvement followed a similar delayed pattern to that of respiratory symptoms, which involved no change at month 1 but significant improvement from month 2 onward, combined with the largest item-level reductions in uncontrollable worry, low mood, and feeling sad. These findings suggest a broad amelioration of emotional distress rather than improvement in isolated symptoms.

One notable finding is the global well-being trajectory, which was the only domain with a positive slope, and the fact that the adjusted improvement was already significant at month 1. This early peak aligns with the prior longitudinal study by TenVergert et al [37], who found significant improvement in physical mobility, pain, energy, and emotional reaction at the first postoperative month. This initial peak may represent a “honeymoon effect”—an initial euphoria from surviving major surgery and receiving a new organ combined with relief from severe preoperative symptoms. However, the LMM estimates suggest a complex pattern wherein well-being peaked at month 4 rather than at month 1, and a nonsignificant dip was observed at month 3. This dip may reflect encounter with postoperative complications such as atrial arrhythmias [38], hyperglycemia [39], acute rejection [40], and medication side effects [41]. Despite this fluctuation, well-being scores remained above preoperative levels at all postoperative time points, reflecting a dynamic trajectory of initial optimism, subsequent adaptation, and long-term net improvement. Unmet life expectations after transplantation may contribute to increased stress [42]. To mitigate the emotional fluctuations observed after month 1, psychological interventions should be tailored to the early recovery phase, potentially beginning during the transition from hospital to home [43,44].

## Comparison With International Studies

The overall improvement trajectory we observed is consistent with that of the original LT-QoL validation cohort from the United States in the study by Singer et al [17], although direct numerical comparisons are limited by differences in follow-up protocols and population characteristics and our use of LMMs rather than raw score comparisons. Our findings also align with those of Ricotti et al [43], who reported significant QoL and functional improvements after LTx in a Chinese single-center study. The cultural specificities of our population, particularly regarding psychological and sexual health reporting, highlight the importance of developing culturally adapted PRO assessment strategies for diverse transplant populations.



XSL•FO
RenderX

## Clinical Implications

Our findings carry several important clinical implications. First, the LMM analysis identified month 1 as a critical window characterized by transient worsening in chest pain and stomach cramps, whereas global well-being simultaneously peaked. This paradoxical pattern characterized by symptom worsening alongside perceived improvement suggests that targeted gastrointestinal symptom management and immunosuppressive protocol adjustment should be implemented during this period without assuming that patients' optimism reflects symptom resolution. Second, the delayed cognitive improvement (significant only from month 4) highlights the need for continued monitoring and rehabilitation throughout the recovery period. Long-term tracking of cognitive function changes is important for capturing early signals of impairment, allowing for timely implementation of targeted interventions such as cognitive training, rehabilitation, and support group therapy [45]. Postoperative rehabilitation, including physical therapy, aerobic exercise, and resistance training, can also help patients adapt to their new life [46]. Third, our findings demonstrate that PRO-based monitoring using disease-specific instruments and LMM analysis can capture clinically meaningful QoL changes that may not be detectable by standard clinical assessments alone. The integration of routine PRO monitoring into posttransplant care can facilitate early identification of patients requiring additional support, enable personalized care planning, and improve shared decision-making between clinicians and patients. In our clinical practice, we have established a peer support program named the LTx Patients Association, which invites well-recovered transplant recipients to provide guidance and emotional support to recent transplant recipients, encouraging social reintegration through community-based activities.

## Limitations

Several limitations should be acknowledged. First, our data were derived from a relatively small cohort at a single center with follow-up limited to 6 months. Although we used LMM, which appropriately handles incomplete data under the MAR assumption, survivorship bias remains a concern in that patients who remained in follow-up may represent a healthier subset, potentially leading to overestimation of QoL improvement at later time points. Second, the reasons for missing data are multifactorial, including patient death, geographic barriers to follow-up, health deterioration preventing participation, and the urgency of transplant surgeries precluding baseline assessment for some patients. While our LMM approach provides valid inference under the MAR assumption and the ICCs (0.25 - 0.41 across domains) confirmed the appropriateness of the mixed-effects framework, unmeasured confounding related to differential attrition cannot be ruled out. Third, while the LT-QoL questionnaire has been validated for the Chinese context, certain domains, particularly sexual function, may be subject to cultural response bias, as discussed above. The high missingness in this domain severely limits statistical power. Additionally, the relatively short follow-up period may not capture long-term QoL fluctuations. Future studies should implement strategies to minimize attrition, establish multicenter cohorts, extend follow-up duration, and develop culturally tailored assessment approaches to comprehensively evaluate QoL in Chinese patients undergoing LTx.

## Conclusions

In this prospective cohort study, LTx significantly improved overall QoL in patients with end-stage lung diseases over 6 months, as demonstrated by random-intercept LMM analysis across 7 QoL domains. The health perception domain showed the earliest and largest improvement, followed by respiratory symptoms, anxiety and depression, and cognitive limitations. Additionally, the first postoperative month represents a critical intervention window for gastrointestinal symptom management and psychological support. These findings support the integration of disease-specific PRO monitoring and LMM-based longitudinal analysis into posttransplant clinical care to identify critical intervention time points, personalize supportive care, and optimize long-term recovery outcomes.

## Acknowledgments

The authors thank all participants who generously contributed to this study. The authors confirm that no generative artificial intelligence (AI) or AI-assisted technologies were used in the research design, data analysis, or manuscript writing. AI tools were used only for language polishing, and all content has been reviewed and approved by the authors.

## Funding

This work was funded by the Shaanxi Province Key Research and Development Program (grant 2024SF-YBXM-083), Shaanxi Province Postdoctoral Science Foundation, and National Natural Science Foundation of China (grant 82470107).

## Data Availability

The data that support the findings of this study are available from the corresponding author on reasonable request.

## Authors' Contributions

PL performed data collection, conducted data analysis, interpreted the findings, applied statistical methods, and drafted the manuscript. JW undertook data collection, executed the data analysis, and prepared the supplementary materials. HW, LW, and SG contributed to data collection and participated in data interpretation. CH engaged in data interpretation. GZ performed



XSL•FO
RenderX

conceptual design and provided supervision. All authors provided critical revision of the manuscript and approved the final version.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures, tables, and details of the lung transplantation management procedure.
[DOCX File, 445 KB - i-jmr_v15i1e86211_app1.docx ]

## References


1. Chambers DC, Perch M, Zuckermann A, et al. The International Thoracic Organ Transplant Registry of the International Society for Heart and Lung Transplantation: thirty-eighth adult lung transplantation report - 2021; focus on recipient characteristics. J Heart Lung Transplant 2021 Oct;40(10):1060-1072. [doi: 10.1016/j.healun.2021.07.021] [Medline: 34446355]
2. Hartert M, Senbaklavacin O, Gohrbandt B, Fischer BM, Buhl R, Vahld CF. Lung transplantation: a treatment option in end-stage lung disease. Dtsch Arztebl Int 2014 Feb 14;111(7):107-116. [doi: 10.3238/arztebl.2014.0107] [Medline: 24622680]
3. Chacon-Alberty L, Fernandez R, Jindra P, et al. Primary graft dysfunction in lung transplantation: a review of mechanisms and future applications. Transplantation 2023 Aug 1;107(8):1687-1697. [doi: 10.1097/TP.0000000000004503] [Medline: 36650643]
4. Jing L, Chen W, Zhao L, et al. Acute kidney injury following adult lung transplantation. Chin Med J (Engl) 2021 Dec 16;135(2):172-180. [doi: 10.1097/CM9.0000000000001636] [Medline: 34924505]
5. Orrego CM, Cordero-Reyes AM, Estep JD, et al. Atrial arrhythmias after lung transplant: underlying mechanisms, risk factors, and prognosis. J Heart Lung Transplant 2014 Jul;33(7):734-740. [doi: 10.1016/j.healun.2014.02.032] [Medline: 24755110]
6. Delbove A, Senage T, Gazengel P, et al. Incidence and risk factors of anastomotic complications after lung transplantation. Ther Adv Respir Dis 2022;16:17534666221110354. [doi: 10.1177/17534666221110354] [Medline: 35894432]
7. Phoompoung P, Villalobos AP, Jain S, Foroutan F, Orchanian-Cheff A, Husain S. Risk factors of invasive fungal infections in lung transplant recipients: a systematic review and meta-analysis. J Heart Lung Transplant 2022 Feb;41(2):255-262. [doi: 10.1016/j.healun.2021.09.014] [Medline: 34872817]
8. Kim HH, Jo KW, Shim TS, et al. Incidence, risk factors, and clinical characteristics of airway complications after lung transplantation. Sci Rep 2023 Jan 12;13(1):667. [doi: 10.1038/s41598-023-27864-1] [Medline: 36635329]
9. Loor G, Mattar A, Schaheen L, Bremner RM. Surgical complications of lung transplantation. Thorac Surg Clin 2022 May;32(2):197-209. [doi: 10.1016/j.thorsurg.2022.01.003] [Medline: 35512938]
10. Miggins JJ, Reul RM Jr, Barrett S, et al. Twenty-year survival following lung transplantation. J Thorac Dis 2023 Jun 30;15(6):2997-3012. [doi: 10.21037/jtd-22-1414] [Medline: 37426158]
11. Gries CJ, Rue TC, Heagerty PJ, Edelman JD, Mulligan MS, Goss CH. Development of a predictive model for long-term survival after lung transplantation and implications for the lung allocation score. J Heart Lung Transplant 2010 Jul;29(7):731-738. [doi: 10.1016/j.healun.2010.02.007] [Medline: 20382034]
12. Ågren S, Sjöberg T, Ekmehag B, Wiborg MB, Ivarsson B. Psychosocial aspects before and up to 2 years after heart or lung transplantation: experience of patients and their next of kin. Clin Transplant 2017 Mar;31(3). [doi: 10.1111/ctr.12905] [Medline: 28039882]
13. Duffy JP, Kao K, Ko CY, et al. Long-term patient outcome and quality of life after liver transplantation: analysis of 20-year survivors. Ann Surg 2010 Oct;252(4):652-661. [doi: 10.1097/SLA.0b013e3181f5f23a] [Medline: 20881772]
14. McAdams-DeMarco MA, Olorundare IO, Ying H, et al. Frailty and postkidney transplant health-related quality of life. Transplantation 2018 Feb;102(2):291-299. [doi: 10.1097/TP.0000000000001943] [Medline: 28885489]
15. Tokuno J, Chen-Yoshikawa TF, Oga T, et al. Analysis of optimal health-related quality of life measures in patients waitlisted for lung transplantation. Can Respir J 2020;2020:4912920. [doi: 10.1155/2020/4912920] [Medline: 32211085]
16. Singer LG, Chowdhury NA, Faughnan ME, et al. Effects of recipient age and diagnosis on health-related quality-of-life benefit of lung transplantation. Am J Respir Crit Care Med 2015 Oct 15;192(8):965-973. [doi: 10.1164/rccm.201501-0126OC] [Medline: 26131729]
17. Singer JP, Soong A, Chen J, et al. Development and preliminary validation of the Lung Transplant Quality of Life (LT-QOL) survey. Am J Respir Crit Care Med 2019 Apr 15;199(8):1008-1019. [doi: 10.1164/rccm.201806-1198OC] [Medline: 30303408]
18. Yu H, Yu Q, Nie Y, et al. Data quality of longitudinally collected patient-reported outcomes after thoracic surgery: comparison of paper- and web-based assessments. J Med Internet Res 2021 Nov 9;23(11):e28915. [doi: 10.2196/28915] [Medline: 34751657]




XSL•FO
RenderX

19. Basch E. Patient-reported outcomes — harnessing patients' voices to improve clinical care. N Engl J Med 2017 Jan 12;376(2):105-108. [doi: 10.1056/NEJMp1611252] [Medline: 28076708]
20. China Food and Drug Regulatory Information Network. China Center for Food and Drug International Exchange. URL: https://www.ccfdie.org/en/index.htm [accessed 2025-02-19]
21. Huo YT, Liu JC, Cao SX, et al. Construction of natural population cohort on telephone follow-up management quality control system and discussion regarding critical issues by REDCap system [Article in Chinese]. Zhonghua Liu Xing Bing Xue Za Zhi 2023 Dec 10;44(12):1970-1976. [doi: 10.3760/cma.j.cn112338-20230306-00125] [Medline: 38129155]
22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
23. Yamane M, Date H, Okazaki M, Toyooka S, Aoe M, Sano Y. Long-term improvement in pulmonary function after living donor lobar lung transplantation. J Heart Lung Transplant 2007 Jul;26(7):687-692. [doi: 10.1016/j.healun.2007.04.008] [Medline: 17613398]
24. Bowdish ME, Pessotto R, Barbers RG, Schenkel FA, Starnes VA, Barr ML. Long-term pulmonary function after living-donor lobar lung transplantation in adults. Ann Thorac Surg 2005 Feb;79(2):418-425. [doi: 10.1016/j.athoracsur.2004.07.008] [Medline: 15680807]
25. Theodore J, Jamieson SW, Burke CM, et al. Physiologic aspects of human heart-lung transplantation. Pulmonary function status of the post-transplanted lung. Chest 1984 Sep;86(3):349-357. [doi: 10.1378/chest.86.3.349] [Medline: 6432455]
26. Bartels MN, Armstrong HF, Gerardo RE, et al. Evaluation of pulmonary function and exercise performance by cardiopulmonary exercise testing before and after lung transplantation. Chest 2011 Dec;140(6):1604-1611. [doi: 10.1378/chest.10-2721] [Medline: 21680643]
27. Singhal R, Shah YM. Oxygen battle in the gut: hypoxia and hypoxia-inducible factors in metabolic and inflammatory responses in the intestine. J Biol Chem 2020 Jul 24;295(30):10493-10505. [doi: 10.1074/jbc.REV120.011188] [Medline: 32503843]
28. Anup R, Balasubramanian KA. Surgical stress and the gastrointestinal tract. J Surg Res 2000 Aug;92(2):291-300. [doi: 10.1006/jsre.2000.5874] [Medline: 10896836]
29. Bulbuloglu S, Gunes H, Saritas S. The effect of long-term immunosuppressive therapy on gastrointestinal symptoms after kidney transplantation. Transpl Immunol 2022 Feb;70:101515. [doi: 10.1016/j.trim.2021.101515] [Medline: 34922024]
30. Chan S, Cao C, Pascoe EM, et al. Patient-reported gastrointestinal symptoms and the association with quality of life following kidney transplantation. Kidney Int Rep 2020;6(1):138-145. [doi: 10.1016/j.ekir.2020.10.013] [Medline: 33426393]
31. Tomasi R, Klemm M, Hinske CL, et al. Impairment of cognitive function in different domains early after lung transplantation. J Clin Psychol Med Settings 2022 Mar;29(1):103-112. [doi: 10.1007/s10880-021-09787-z] [Medline: 34009540]
32. Cao L, Ye S, Chen Y, Pei Y, Chen J, Li X. Longitudinal study on the trajectory and influencing factors of cognitive dysfunction in lung transplantation patients. Transpl Immunol 2024 Jun;84:102053. [doi: 10.1016/j.trim.2024.102053] [Medline: 38750974]
33. Smith PJ, Blumenthal JA, Hoffman BM, Davis RD, Palmer SM. Postoperative cognitive dysfunction and mortality following lung transplantation. Am J Transplant 2018 Mar;18(3):696-703. [doi: 10.1111/ajt.14570] [Medline: 29087035]
34. Yang LS, Shan LL, Saxena A, Morris DL. Liver transplantation: a systematic review of long-term quality of life. Liver Int 2014 Oct;34(9):1298-1313. [doi: 10.1111/liv.12553] [Medline: 24703371]
35. Muehrer RJ. Sexuality, an important component of the quality of life of the kidney transplant recipient. Transplant Rev (Orlando) 2009 Oct;23(4):214-223. [doi: 10.1016/j.trre.2009.06.001] [Medline: 19647423]
36. Kaplan A, Korenjak M, Brown RS. Post-liver transplantation patient experience. J Hepatol 2023 Jun;78(6):1234-1244. [doi: 10.1016/j.jhep.2023.01.008] [Medline: 37208108]
37. TenVergert EM, Essink-Bot ML, Geertsma A, van Enckevort PJ, de Boer WJ, van der Bij W. The effect of lung transplantation on health-related quality of life: a longitudinal study. Chest 1998 Feb;113(2):358-364. [doi: 10.1378/chest.113.2.358] [Medline: 9498952]
38. Marazzato J, Eikermann M, Di Biase L. Management of atrial arrhythmias after lung transplant. JACC Clin Electrophysiol 2023 Aug;9(8 Pt 3):1824-1835. [doi: 10.1016/j.jacep.2023.01.021] [Medline: 37648342]
39. Weir W, Han MK, Shih TY, et al. Effects of hyperglycemia on primary graft dysfunction in lung transplantation. J Heart Lung Transplant 2013 Apr;32(4):S271. [doi: 10.1016/j.healun.2013.01.709]
40. Todd JL, Neely ML, Kopetskie H, et al. Risk factors for acute rejection in the first year after lung transplant. A multicenter study. Am J Respir Crit Care Med 2020 Aug 15;202(4):576-585. [doi: 10.1164/rccm.201910-1915OC] [Medline: 32379979]
41. Lung transplant medicines - organ transplantation. National Health Service. URL: https://www.nhsbt.nhs.uk/organ-transplantation/lung/living-with-a-lung-transplant/lung-transplant-medicines/ [accessed 2025-02-10]
42. Holzner B, Kemmler G, Kopp M, et al. Preoperative expectations and postoperative quality of life in liver transplant survivors. Arch Phys Med Rehabil 2001 Jan;82(1):73-79. [doi: 10.1053/apmr.2001.19013] [Medline: 11239289]
43. Ricotti S, Martinelli V, Caspani P, et al. Changes in quality of life and functional capacity after lung transplantation: a single-center experience. Monaldi Arch Chest Dis 2017 Dec 18;87(3):831. [doi: 10.4081/monaldi.2017.831]




XSL•FO
RenderX

44. Vermeulen KM, Ouwens JP, van der Bij W, de Boer WJ, Koëter GH, TenVergert EM. Long-term quality of life in patients surviving at least 55 months after lung transplantation. Gen Hosp Psychiatry 2003;25(2):95-102. [doi: 10.1016/s0163-8343(03)00007-0] [Medline: 12676422]
45. Castle A. A guide to interventions for cognitive impairment. HappyNeuron Blog. URL: https://news.happyneuronpro.com/interventions-for-cognitive-impairment/ [accessed 2025-02-10]
46. Wickerson L, Rozenberg D, Janaudis-Ferreira T, et al. Physical rehabilitation for lung transplant candidates and recipients: an evidence-informed clinical approach. World J Transplant 2016 Sep 24;6(3):517-531. [doi: 10.5500/wjt.v6.i3.517] [Medline: 27683630]

## Abbreviations

**ICC:** intraclass correlation coefficient
**LMM:** linear mixed model
**LRT:** likelihood ratio test
**LT-QoL:** Lung Transplant Quality of Life
**LTx:** lung transplantation
**MAR:** missing at random
**MD:** mean difference
**PRO:** patient-reported outcome
**QoL:** quality of life
**REDCap:** Research Electronic Data Capture

*Edited by M Balcarras; submitted 20.Oct.2025; peer-reviewed by A AL-Asadi, C Lei, H Yu; revised version received 10.Apr.2026; accepted 30.Apr.2026; published 10.Aug.2026.*

*Please cite as:*
*Li P, Wang J, Wang H, Li S, Wang L, Gao S, Hu C, Zhang G*
*Patient-Reported Quality of Life in Lung Transplantation: Prospective Cohort Study*
*Interact J Med Res 2026;15:e86211*
*URL: https://www.i-jmr.org/2026/1/e86211*
*doi:10.2196/86211*

© Peng Li, Jiawei Wang, Hongyi Wang, Shuo Li, Lei Wang, Shan Gao, Chunxiao Hu, Guangjian Zhang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 10.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Cardiac Resynchronization Therapy With Defibrillator Using the JROAD-DPC Database: Cost-Effectiveness Analysis

Kazuki Ohashi[1*], PhD; Masaya Watanabe[2,3*], MD, PhD; Yasuhiro Morii[1,4], MBA, PhD; Hisashi Yokoshiki[5], MD, PhD; Kengo Kusano[6], MD, PhD; Katsuhiko Imai[7,8], MD, PhD; Masahiko Takagi[9], MD, PhD; Teiichi Yamane[10], MD, PhD; Hiroshi Tada[11], MD, PhD; Katsuhiko Ogasawara[1,12], MBA, PhD

[1]Faculty of Health Sciences, Hokkaido University, N12-W5, Kita-ku, Sapporo, Hokkaido, Japan

[10]Department of Cardiology, Jikei University School of Medicine, Tokyo Minato-ku, Tokyo, Japan

[11]Department of Cardiovascular Medicine, Faculty of Medical Sciences, University of Fukui, Fukui, Fukui, Japan

[12]Faculty of Engineering, Muroran Institute of Technology, Muroran, Hokkaido, Japan

[2]Department of Cardiovascular Medicine, Caress Memorial Hospital, Sapporo, Hokkaido, Japan

[3]Department of Cardiovascular Medicine, Graduate School of Medicine, Hokkaido University, Sapporo, Hokkaido, Japan

[4]Center for Outcomes Research and Economic Evaluation for Health, National Institute of Public Health, Wako, Saitama, Japan

[5]Department of Cardiovascular Medicine, Sapporo City General Hospital, Sapporo, Hokkaido, Japan

[6]Department of Cardiovascular Medicine, National Cerebral and Cardiovascular Center, Suita, Osaka, Japan

[7]Department of Cardiovascular Surgery, Kure Medical Center, Kure, Hiroshima, Japan

[8]Department of Cardiovascular Surgery, Hiroshima University, Hiroshima, Hiroshima, Japan

[9]Division of Cardiac Arrhythmia, Kansai Medical University Medical Centre, Moriguchi, Osaka, Japan

[*]these authors contributed equally

**Corresponding Author:**
Katsuhiko Ogasawara, MBA, PhD
Faculty of Health Sciences, Hokkaido University, N12-W5, Kita-ku, Sapporo, Hokkaido, Japan

## Abstract

**Background:** Cardiac resynchronization therapy with defibrillator (CRT-D) improves survival, reduces hospitalization, and enhances quality of life in patients with heart failure and reduced ejection fraction (HFrEF). As heart failure prevalence increases in aging societies such as Japan, the associated clinical and economic burden continues to rise. Previous cost-effectiveness analyses conducted in multiple countries indicate that CRT-D may be cost-effective in selected patients with HFrEF. However, its cost-effectiveness within the Japanese health care system remains uncertain.

**Objective:** This study aimed to evaluate the cost-effectiveness of CRT-D in patients with HFrEF within the Japanese health care setting.

**Methods:** A partitioned survival model was developed with 3 health states: after treatment (follow-up), hospitalization, and death. Survival for CRT-D was estimated by reconstructing individual patient-level data from the Kaplan-Meier curve of the RAFT (Resynchronization-Defibrillation for Ambulatory Heart Failure Trial) study using the method proposed by Guyot et al followed by fitting multiple parametric models; the gamma distribution was selected for the base case analysis. Survival for optimal medical therapy (OMT), the comparator, was estimated by applying a hazard ratio from a published meta-analysis. Hospitalization rates and device longevity were derived from prior studies. Cost estimates were obtained from the JROAD-DPC (Japanese Registry Of All cardiac and vascular Disease-Diagnostic Procedure Combination) database and the Japanese medical fee schedule. Utility values were assigned according to New York Heart Association class assuming treatment-specific distributions. The analysis was conducted from the public health care payer perspective using a monthly cycle over a 20-year time horizon. Deterministic and probabilistic sensitivity analyses were performed. Additionally, scenario analyses were conducted by varying the duration of treatment effect.

**Results:** In the base case analysis, per capita costs were ¥12,258,410 (¥1=US $0.006 as of July 7, 2026) for CRT-D and ¥640,056.90 for OMT, resulting in an incremental cost of ¥11,618,353. CRT-D generated 7.07 quality-adjusted life years (QALYs) compared with 4.75 QALYs for OMT, yielding an incremental gain of 2.32 QALYs. The incremental cost-effectiveness ratio (ICER) was ¥5,009,880 per QALY. Scenario analyses showed that, when treatment effect waned after 7.5 years, the ICER increased to ¥5,423,235 per QALY. When the time horizon was shortened to 10 years or extended to 30 years, the ICERs were ¥8,523,072 and ¥4,386,803 per QALY, respectively. Deterministic sensitivity analysis identified CRT-D efficacy (hazard ratio), discount rate, and initial treatment cost as primary ICER drivers. Probabilistic sensitivity analysis produced a median ICER of



XSL•FO
RenderX

¥5,022,618 (IQR ¥4,448,306–¥5,760,425) per QALY, with a 95% credible interval of ¥3,804,418 to ¥7,178,795. At a willingness-to-pay value of ¥5,000,000 per QALY, CRT-D had a 49.2% probability of being cost-effective.

**Conclusions:** CRT-D demonstrated acceptable cost-effectiveness in patients with HFrEF in Japan. Treatment efficacy and initial cost were the primary determinants of economic value, emphasizing the importance of appropriate patient selection and strategies to reduce device costs.

***(Interact J Med Res 2026;15:e94073)*** doi:10.2196/94073

**KEYWORDS**

cost-effectiveness analysis; cardiac resynchronization therapy with defibrillator; heart failure and reduced ejection fraction; JROAD-DPC database; Japan

## Introduction

Cardiac resynchronization therapy with defibrillator (CRT-D) is an established device-based treatment for patients with heart failure (HF) and reduced ejection fraction (HFrEF) [1]. By integrating biventricular pacing with defibrillation capability, CRT-D reduces all-cause mortality, prevents sudden cardiac death, decreases hospitalization for HF, and improves functional status and quality of life in appropriately selected patients [2,3]. Population aging in high-income countries, including Japan, has accelerated the increase in HF prevalence, resulting in a growing population eligible for CRT-D implantation [4,5]. In Japan, earlier reports estimated HF prevalence at approximately 1% [6,7]; however, recent analyses using large administrative claims databases have reported substantially higher estimates of 3% to 6.5% [8]. The European Society of Cardiology estimates that HF accounts for up to 2% of total health care expenditures in high-income countries [9]. These findings indicate a rising HF burden with significant implications for health care spending and resource allocation.

Given constrained health care resources, prioritizing cost-effective therapies is essential. Cost-effectiveness studies of CRT-D conducted outside Japan have produced mixed results, and its economic value remains controversial. Existing evidence indicates that cost-effectiveness varies according to clinical characteristics, including QRS duration, left bundle branch block (LBBB) status, and ischemic vs nonischemic etiology [10-12]. CRT-D has demonstrated cost-effectiveness in patients with QRS duration of 150 ms or higher and New York Heart Association (NYHA) class II to III symptoms, whereas in patients with more advanced HF, such as NYHA class IV, cardiac resynchronization therapy with pacemaker may represent a more economically favorable option [11]. However, most evaluations have been conducted in Western health care systems, limiting generalizability to Japan because of differences in reimbursement structures and cost frameworks. In Japan, cost-effectiveness studies of CRT-D remain limited, and existing health technology assessments have primarily evaluated incremental device features such as automated optimization algorithms and reactive atrial-based antitachycardia pacing (rATP) [13,14]. Therefore, this study aimed to evaluate the cost-effectiveness of CRT-D in the Japanese health care setting using large-scale registry data to provide real-world evidence supporting clinical and policy decision-making.

## Methods

### Overview

The analysis was conducted from the public health care payer perspective using a monthly cycle over a 20-year time horizon, consistent with Japanese guidelines [15]. A reference value of ¥5,000,000 (¥1=US $0.006 as of July 7, 2026) per quality-adjusted life year (QALY) was used to evaluate cost-effectiveness. The incremental cost-effectiveness ratio (ICER) of CRT-D was assessed relative to optimal medical therapy (OMT).

### Study Population

The modeled cohort reflected the CRT-D group of the RAFT (Resynchronization-Defibrillation for Ambulatory Heart Failure Trial) study [16]. The mean age was 66.3 (SD 9.3) years, and 84.8% of the participants were male. NYHA functional class distribution was 76.7% class II and 23.3% class III. The mean left ventricular ejection fraction was 22.5% (SD 5.4%), and mean QRS duration was 158.1 (SD 23.8) ms. Ischemic etiology accounted for 68.7% of cases, and LBBB morphology was present in 71.7%.

### Model Structure and Inputs

A partitioned survival model was developed to evaluate CRT-D cost-effectiveness in patients with HFrEF, with OMT as the comparator. The model included 3 health states: after treatment (follow-up), hospitalization, and death. Hospitalization events included admissions due to cardiac arrest (CA), worsening HF, and CRT-D–related complications. The CRT-D survival curve was derived by digitizing Kaplan-Meier data from long-term RAFT follow-up [16] and reconstructing individual patient-level survival data using the method proposed by Guyot et al [17]. Parametric survival models, including exponential, Weibull, Gompertz, log-logistic, log-normal, gamma, and generalized gamma distributions, were evaluated. On the basis of the Akaike information criterion (AIC), Bayesian information criterion (BIC), and visual inspection, the gamma distribution provided the best fit and was selected. Comparative AIC and BIC results can be found in Table S1 in Multimedia Appendix 1. Survival for OMT was estimated by applying a hazard ratio (HR) for all-cause mortality from the meta-analysis by Woods et al [3] (0.54, 95% CI 0.43-0.69), assuming proportional hazards.

The primary outcome was QALYs, integrating survival and health-related quality of life. Utility values were assigned according to NYHA functional class and differed by treatment



XSL•FO
RenderX

group. Adverse outcomes were modeled as hospitalization events, including HF, CA, device-related complications, and temporary disutility associated with device replacement. The incidence of HF hospitalization was obtained from a prior study [18], which reported cumulative rates of 19.5% in the CRT-D group and 26.1% in the OMT group over a median follow-up of 40 months. These cumulative incidences were converted into monthly transition probabilities for each treatment group, assuming a constant hazard throughout the observation period. The incidence of hospitalization due to CA was estimated from indirect published evidence [19-21]. In the CRT-D group, Tseng et al [19] reported that 6.4% of all-cause deaths following cardiac implantable electronic device therapy were attributable to CA. On the basis of the annual all-cause mortality reported in the MADIT-RIT (Multicenter Automatic Defibrillator Implantation Trial–Reduce Inappropriate Therapy) trial, this proportion corresponded to an estimated CA incidence of 0.14% per year in the CRT-D group. In the OMT group, the annual incidence of CA was derived from a previously reported rate of sudden cardiac death in the conventional therapy group (4.5% per year) [21], used as a surrogate marker for arrhythmic CA requiring hospitalization. This yielded an estimated annual incidence of 6.51%. Device-related complications were modeled exclusively in the CRT-D group during follow-up, with a cumulative incidence of 14% over 48 months [22]. This value was converted into a monthly transition probability assuming a constant hazard. Among previously reported complications [23,24], superficial and systemic infections accounted for 21.7% and 26.6%, respectively, whereas lead dislodgement accounted for 51.7%. Device replacement was assumed in cases of systemic infection or lead dislodgement. Device longevity was assumed to be 111 months based on a recent large-scale observational report from a single manufacturer [25]. In the model, device replacements due to battery depletion were assumed to recur continuously without upper limits as long as the simulated cohort survived.

We included costs associated with the initial intervention; device replacement; hospitalizations (CA, HF, and device-related complications); and additional follow-up, including remote monitoring and inappropriate shocks. Costs related to procedure-related complications at initial implantation were incorporated into the initial intervention cost. Cost inputs were derived from analyses of the JROAD-DPC database (approval 20240001) and the Japanese medical fee schedule, which provides standardized reimbursement for each medical service [26]. The database, administered by the Japanese Circulation Society, contains diagnosis procedure combination and per-diem payment system data from approximately 1500 hospitals as of 2024 and covers 2012 to 2024 [27,28]. It represents one of the largest nationwide databases for cardiac and vascular diseases in Japan. For the cost analysis in this study, we extracted data on Japanese patients who underwent initial CRT-D implantation or device replacement between 2012 and 2022. Cost extraction was based on diagnostic and procedure codes. CRT-D implantation and device replacement were identified using procedure codes in the claims database. Hospitalization costs for HF and CA were assumed to be identical between groups as they reflect baseline disease management. Clinical cost differences between groups were captured by adding device-related complications to the CRT-D group. Device-related complications were classified as superficial or systemic infections using *International Classification of Diseases, 10th Revision*, diagnostic codes. The specific codes used for each definition are provided in Table S2 in Multimedia Appendix 1. Utilities were assigned according to NYHA functional class (I-IV), and cohort-level utility was calculated as a weighted average based on the distribution of NYHA classes [29]. Because treatment was assumed to modify the NYHA class distribution, different stable-state utility values were applied to each treatment group. Specifically, CRT-D was assumed to improve NYHA functional class compared with OMT. On the basis of the COMPANION (Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure) trial [30], the proportion of patients with NYHA class improvement at 6 months was 57% in the CRT-D group and 38% in the OMT group. These improvement rates were applied to the baseline NYHA class distribution to estimate follow-up utility. The magnitude of improvement was assumed to be equivalent for patients with baseline NYHA class II and III, and the difference in improvement rates between treatment groups was assumed to persist throughout the time horizon. On the basis of these assumptions, the mean health-related utility during follow-up was 0.805 for CRT-D and 0.781 for OMT. For hospitalization events, mean utility values were 0.649 for HF hospitalization and 0.44 for CA hospitalization. Disutility values of −0.10 for complications and –0.05 for device replacement were applied [31]. All model inputs are summarized in Table 1. Costs and utilities were discounted at an annual rate of 2%, consistent with Japanese cost-effectiveness analysis guidelines [15]. All analyses were conducted using R (version 4.4.3; R Foundation for Statistical Computing).



XSL•FO
RenderX

**Table .** Summary of model inputs (¥1=US $0.006 as of July 7, 2026).

| Parameter and group | Values | Distribution | Source |
|---|---|---|---|
| Hazard ratio (95% CI)[a] | 0.54 (0.43 to 0.69) | Log-normal | [3] |
| Events rate per month (%) | | Beta | |
| HF[b] | | | |
| CRT-D[c] | 0.54 (0.46 to 0.62) | | [19] |
| OMT[d] | 0.75 (0.66 to 0.86) | | [19] |
| CA[e] | | | |
| CRT-D | 0.01[f] | | [19,20] |
| OMT | 0.38 (0.22 to 0.67) | | [21] |
| Complication—CRT-D | 0.31 (0.27 to 0.36) | | Assumption |
| Cost (¥) | | Gamma | |
| Initial treatment—CRT-D | 6,607,498 (5,469,735 to 7,745,261) | | DPC[g] |
| Device replacement—CRT-D | 4,817,601 (3,603,617 to 6,031,585) | | DPC |
| HF—Both | 795,777 (636,622 to 954,932) | | DPC |
| CA—Both | 719,335 (575,468 to 863,202) | | DPC |
| Complication—CRT-D | 6,296,972 (5,037,577 to 7,556,366) | | Assumption |
| Additional follow-up—CRT-D | 4900 (–20% to +20%) | | Assumption |
| Utility | | Beta | |
| Follow-up | | | |
| CRT-D | 0.805 (0.792 to 0.820) | | [30] |
| OMT | 0.781 | | [30] |
| HF—Both | 0.649 (–20% to +20%) | | Assumption |
| CA—Both | 0.44 (–20% to +20%) | | Assumption |
| Complication—CRT-D | –0.1 | | [31] |
| Replacement—CRT-D | –0.05 | | [31] |
| Death—Both | 0 | | —[h] |
| Others | | | |
| Device longevity (mo) | | | |
| CRT-D | 111 (95 to 126) | Triangular | [25] |
| Discount rate (%) | | | |
| Both | 2 (0 to 4) | Beta | [15] |

[a]Unless otherwise specified, the lower and upper values represent the parameter bounds used for the sensitivity analysis and do not represent 95% CIs or 95% credible intervals.

[b]HF: heart failure.

[c]CRT-D: cardiac resynchronization therapy with defibrillator.

[d]OMT: optimal medical therapy.

[e]CA: cardiac arrest.

[f]Unless otherwise specified, the lower and upper values represent the parameter bounds used for sensitivity analyses and do not represent 95% CIs or 95% credible intervals. Confidence or credible intervals are not applicable to fixed model parameters.

[g]DPC: Diagnostic Procedure Combination.

[h]Not applicable.



XSL•FO
RenderX

### Scenario and Sensitivity Analysis

Formal subgroup analyses were not performed because the individual patient-level clinical characteristics required to define subgroups were unavailable in the data sources. Instead, heterogeneity in treatment effects was evaluated through sensitivity analyses by varying key parameters, including the HR of CRT-D, across plausible ranges reported in the literature.

Scenario analyses examined uncertainty related to treatment effect duration and time horizon. In one scenario, the treatment effect remained constant for 7.5 years and then declined linearly up to 20 years, consistent with prior studies [12,32]. Specifically, this decline was modeled by allowing the HR to approach 1.0 and the difference in utility values between the groups to approach 0. In separate scenarios, the time horizon was set to 10 and 30 years, whereas the treatment effect remained constant and identical to the base case, which used a 20-year horizon.

To evaluate parameter-level and overall model uncertainty, both deterministic sensitivity analysis (DSA) and probabilistic sensitivity analysis (PSA) were performed. Parameter ranges and probability distributions were applied from published sources when available. For parameters lacking reported ranges or distributions, a –20% to +20% variation from base case values was assumed. The discount rate was varied between 0% and 4% in accordance with relevant guidelines [15].

### Ethical Considerations

This study was reviewed and approved by the institutional review board of the Japanese Heart Rhythm Society (approved 20240001) This study used anonymized data from the JROAD-DPC registry. The requirement for individual informed consent was waived by the Institutional Review Board, and participants were provided with the opportunity to opt out in accordance with the applicable ethical guidelines.

## Results

### Base Case and Scenario Analysis

In the base case analysis, per capita costs were ¥12,258,410 for CRT-D and ¥640,056.90 for OMT, yielding an incremental cost of ¥11,618,353. CRT-D produced 7.07 QALYs compared with 4.75 QALYs for OMT, corresponding to an incremental gain of 2.32 QALYs. The resulting ICER was ¥5,009,880 per QALY (Table 2).

**Table .** Results of base case and scenario analyses (¥1=US $0.006 as of July 7, 2026).

| | Total cost (¥) | Incremental cost (¥) | Total QALYs[a] | Incremental QALYs | ICER[b] (cost per QALY) |
|---|---|---|---|---|---|
| Base case analysis | | | | | |
| CRT-D[c] | 12,258,410 | 11,618,353 | 7.07 | 2.32 | 5,009,880 |
| OMT[d] | 640,056.90 | —[e] | 4.75 | — | — |
| Scenario analysis[f] | | | | | |
| CRT-D | 12,258,410 | 11,597,980 | 7.04 | 2.14 | 5,423,235 |
| OMT | 660,429.90 | — | 4.9 | — | — |

[a]QALY: quality-adjusted life year.

[b]ICER: incremental cost-effectiveness ratio.

[c]CRT-D: cardiac resynchronization therapy with defibrillator.

[d]OMT: optimal medical therapy.

[e]Not applicable.

[f]The treatment effect was assumed to remain constant for 7.5 years and then decline linearly up to 20 years.

In the scenario analysis, the ICER increased to ¥5,423,235 per QALY when the treatment effect was assumed to decline after 7.5 years (Table 2). When the time horizon was shortened to 10 years, the ICER increased to ¥8,523,072 per QALY. Conversely, extending the time horizon to 30 years reduced the ICER to ¥4,386,803 per QALY.

### Sensitivity Analysis

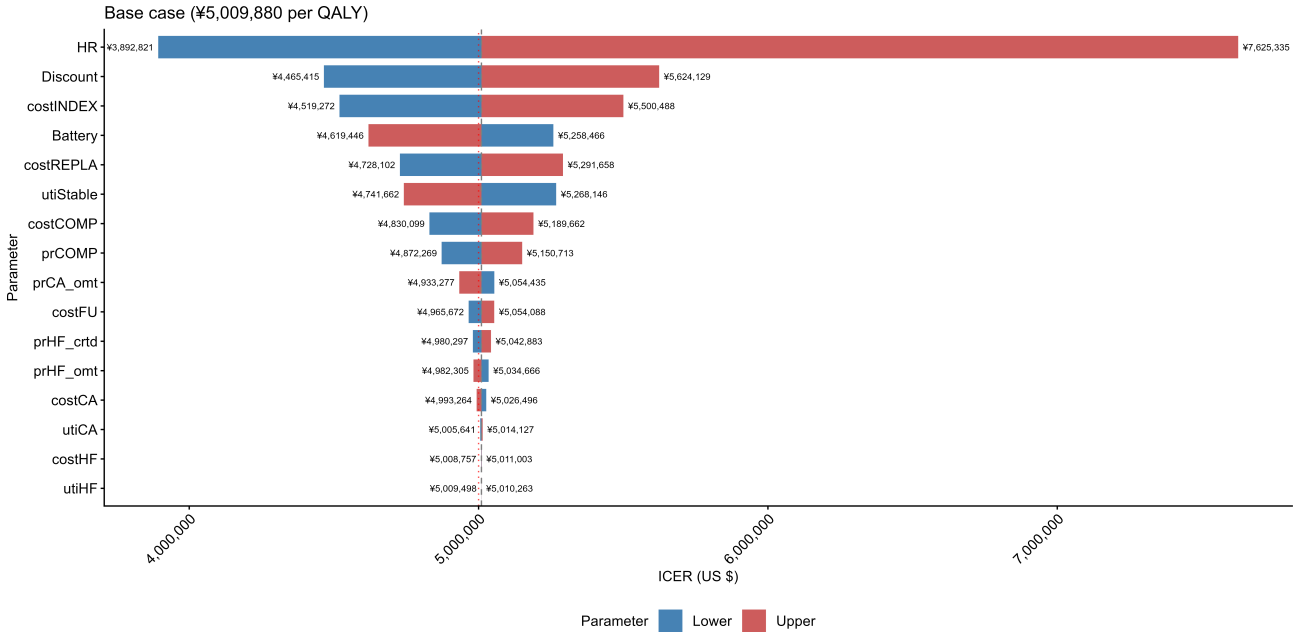
DSA identified the primary drivers of the ICER as CRT-D efficacy (HR), discount rate, and initial treatment cost (“costINDEX”; Figure 1). Variation in CRT-D efficacy produced ICERs ranging from ¥3,892,821 to ¥7,625,335 per QALY. Variation in the discount rate resulted in ICERs ranging from ¥4,465,415 to ¥5,624,129 per QALY, whereas variation in initial treatment cost yielded ICERs ranging from ¥4,519,272 to ¥5,500,488 per QALY.

PSA results are shown in Figure 2. The median ICER was ¥5,022,618 per QALY, with a range of ¥3,804,418 to ¥7,178,795 per QALY across the 2.5th and 97.5th percentiles. At a willingness-to-pay threshold of ¥5,000,000 per QALY, the probability that CRT-D was cost-effective was 49.2%.



XSL•FO
RenderX

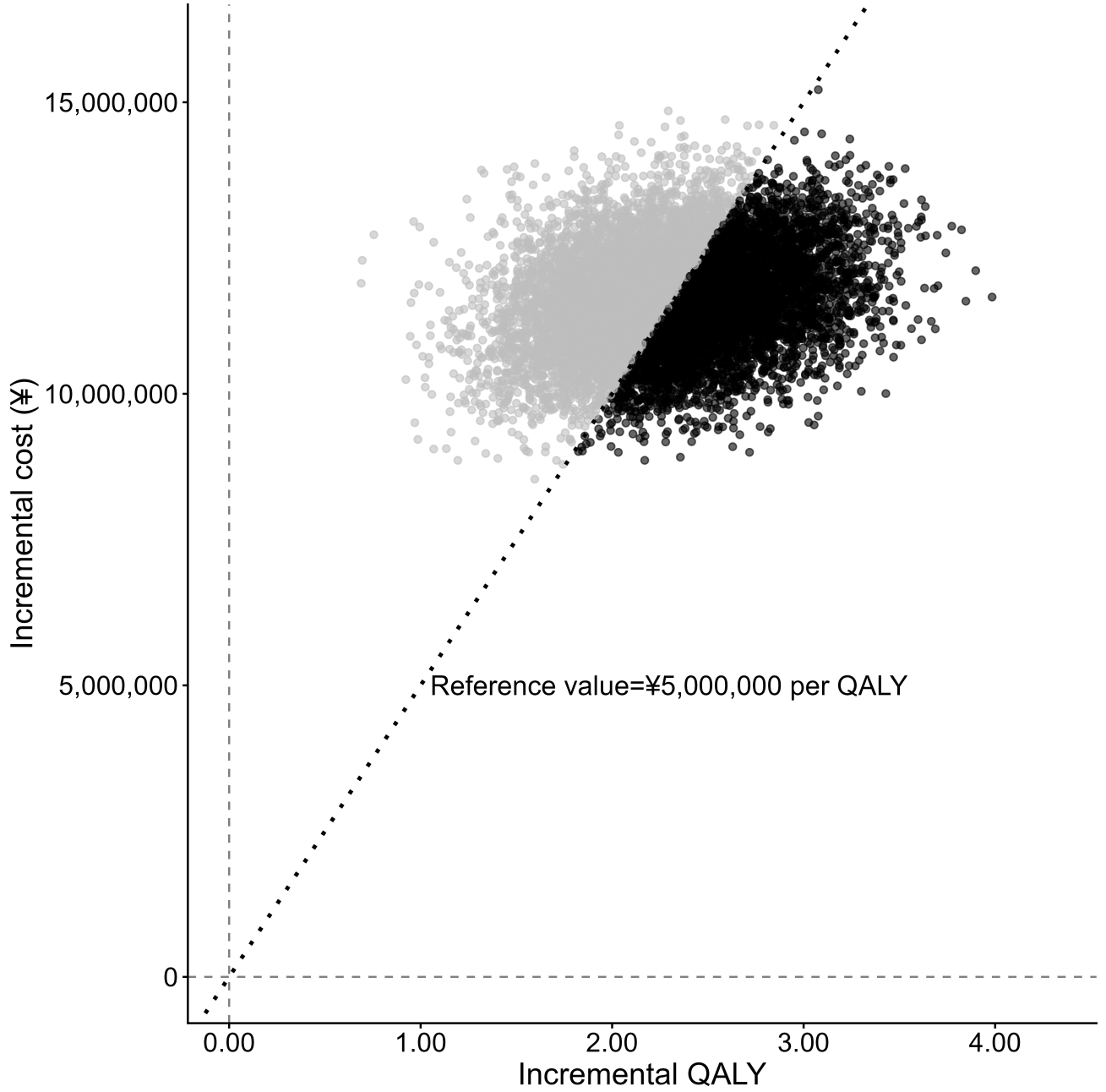
**Figure 1.** Tornado diagram of the deterministic sensitivity analysis. This diagram depicts the effect of varying individual parameters on the incremental cost-effectiveness ratio (ICER). Bar width represents the ICER range obtained when each parameter varied between its lower and upper bounds. The vertical dotted line denotes the base case ICER of ¥5,009,880 (¥1=US $0.006 as of July 7, 2026) per quality-adjusted life year (QALY). Parameters are ordered from top to bottom according to their relative influence on the ICER. costCA: cost of hospitalization for cardiac arrest; costCOMP: cost of hospitalization for complications; costFU: additional follow-up cost; costHF: cost of hospitalization for heart failure; costINDEX: initial treatment cost; costREPLA: device replacement cost; HR: hazard ratio; prCA_crtd: probability of cardiac arrest in the CRT-D group; prCA_omt: probability of cardiac arrest in the OMT group; prCOMP: probability of hospitalization for complications; prHF_crtd: probability of heart failure in the CRT-D group; prHF_omt: probability of heart failure in the OMT group; utiCA: utility of the cardiac arrest state; utiHF: utility of the heart failure state; utiStable: utility of the stable state.






XSL•FO
RenderX

**Figure 2.** Probabilistic sensitivity analysis scatterplot. This figure presents incremental costs vs incremental quality-adjusted life years (QALYs) across 10,000 simulations. The dotted line indicates the willingness-to-pay threshold of ¥5,000,000 per QALY (¥1=US $0.006 as of July 7, 2026). Black dots represent simulations below the reference value (cost-effective), whereas gray dots indicate simulations exceeding the threshold.




## Discussion

### Principal Findings

Using the Japanese JROAD-DPC database, we assessed the cost-effectiveness of CRT-D compared with OMT from a public health insurance perspective. Over a 20-year horizon, CRT-D generated an incremental cost of ¥11,618,353 and a QALY gain of 2.32, yielding an ICER of ¥5,009,880 per QALY. This estimate approximated the Japanese reference value. When a waning treatment effect was incorporated, the ICER increased to ¥5,349,656 per QALY, remaining comparable to the reference threshold. In UK settings, Mealing et al [11] reported that CRT-D was cost-effective in only 10 of 24 patient subgroups at a £30,000 per QALY threshold (£1=US $1.34 as of July 7, 2026), highlighting substantial heterogeneity. In 4 subgroups similar to our population (QRS duration of ≥150 ms and NYHA class II and III with or without LBBB), the ICER for CRT-D vs OMT ranged from £15,000 to £20,000, below the accepted UK threshold of £30,000. In a US study [12] using a reference value of US $100,000 per QALY, CRT-D was consistently the most cost-effective strategy among patients with QRS duration above 150 ms regardless of LBBB status or NYHA class. Although reference thresholds differ across the United Kingdom, United States, and Japan, our findings appear conservative in this international context. An ICER of approximately ¥5 million per QALY suggests that further gains in cost-effectiveness will likely depend on refined patient selection.

DSA identified populations in which higher economic value may be achieved. The principal ICER driver was the HR, indicating that greater treatment effects improve



XSL•FO
RenderX

cost-effectiveness. A meta-analysis by Woods et al [3] showed that, compared with OMT, the HR for CRT-D was lower in female than in male patients irrespective of QRS duration or age. The same analysis demonstrated that, when CRT-D was compared with OMT, the HR approached 1.0 among patients without LBBB or with QRS duration below 150 ms, indicating attenuation of survival benefit in these subgroups. Because survival was modeled using all-cause mortality, CRT-D is likely to be more cost-effective in patients with a lower burden of high-mortality comorbidities. This interpretation aligns with a recent patient-level meta-analysis of 8 randomized trials [33], which reported substantial survival benefit in patients with no comorbidities (adjusted HR 0.54) but diminished benefit in those with 3 or more comorbidities (adjusted HR 0.83). Accordingly, patients with a lower comorbidity burden may tend to exhibit a more favorable ICER, although this trend requires careful interpretation regarding cumulative longitudinal costs. However, a detailed analysis of specific comorbidities was not feasible in the present study, and further research is warranted to evaluate cost-effectiveness stratified by comorbidity profiles. Additionally, the crucial trials underpinning our efficacy estimates were conducted over a decade ago. Since then, contemporary OMT has advanced significantly with modern foundational therapies (ie, angiotensin receptor–neprilysin inhibitors and sodium-glucose cotransporter 2 [SGLT2] inhibitors) [34,35]. In particular, SGLT2 inhibitors are now widely used in Japan as a key component of guideline-directed medical therapy [36]. If these modern drugs substantially improve baseline outcomes, the incremental benefit of CRT-D over OMT might be attenuated in current practice, potentially increasing the ICER. Future modeling studies incorporating clinical data from the modern therapy era are warranted. Recent incremental innovations in CRT-D, including rATP and optimization algorithms, have improved efficacy but increased device costs. Several studies have shown these enhancements to be cost-effective in Japanese and other national settings [13,14,37]. Overall, the additional costs associated with these functional upgrades appear acceptable from a cost-effectiveness perspective. Another influential determinant was the initial treatment cost, driven largely by device price. The current reimbursement for CRT-D devices in Japan is between ¥3,090,000 and ¥4,750,000 depending on the model. Together with lead costs, this accounts for more than half of the initial treatment cost. Device price also affects downstream costs related to battery depletion and infection requiring replacement; therefore, reductions in device price would likely generate substantial savings. In PSA, the 95% credible interval for the ICER was ¥3,804,418 to ¥7,178,795, with a 49.2% probability of remaining below ¥5 million per QALY. These findings indicate moderate parameter uncertainty, with ICER estimates distributed within approximately –¥2.5 million to +¥2.5 million of the reference value.

## Limitations

Several limitations of this study should be acknowledged. First, although the JROAD-DPC database captures a substantial proportion of acute care hospitals in Japan, it does not include all medical institutions. Furthermore, the JROAD-DPC database was not used to derive clinical event probabilities because this database possesses inherent limitations in tracking long-term, longitudinal patient outcomes across different medical institutions once patients are discharged to outpatient care. Consequently, cost inputs may be subject to selection bias, limiting generalizability to all patients with HF in Japan. Second, the 20-year time horizon required extrapolation of survival and costs beyond observed follow-up. To mitigate this limitation, survival curves were generated using parametric modeling of long-term data from the RAFT study. Although using overseas trial data (RAFT) requires caution regarding generalizability, the trial's mean age (66.3 years) is comparable to that of major Japanese CRT-D registries (66.5 or 68.4 years) [38]. Nevertheless, this population-level cost-effectiveness model does not incorporate individual risk factors. Future Japanese studies incorporating patient-level subgroup analyses are needed to refine estimates and support value-based allocation within the health care system. Finally, detailed subgroup analyses by clinical characteristics, such as QRS duration, LBBB status, or comorbidity burden, were not feasible because of data constraints. Instead, the potential influence of these factors was examined through DSA by varying HRs rather than directly estimating subgroup-specific effects.

Despite these limitations, to our knowledge, this study represents one of the first comprehensive cost-effectiveness evaluations of CRT-D in the Japanese health care setting using real-world cost data.

## Conclusions

CRT-D for patients with HFrEF in Japan demonstrated cost-effectiveness within an acceptable range relative to the reference value. Sensitivity analyses identified the HR and initial treatment cost as key determinants of value. These findings emphasize the importance of identifying patient subgroups with a high expected treatment response. Specifically, future cost-effectiveness analyses stratified by comorbidity profiles will provide further valuable insights. From a policy perspective, reducing device costs may further enhance the societal value of this intervention.

## Acknowledgments

The authors would like to thank Editage for English-language editing. During the preparation of this manuscript, the authors used Google Gemini solely for language translation. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the final contents of the publication.



XSL•FO
RenderX

## Funding

The authors declared no financial support was received for this work.

## Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

## Authors' Contributions

Conceptualization: K Ohashi, MW, YM, HY, K Ogasawara
Data curation: K Ohashi, MW, KK, KI, MT, TY, HT
Formal analysis: K Ohashi, MW, YM
Investigation: K Ohashi, MW, YM, HY
Methodology: K Ohashi, MW, YM, HY
Project administration: MW, HY
Supervision: HY, KK, KI, MT, TY, HT, K Ogasawara
Visualization: K Ohashi
Writing—original draft: K Ohashi, MW
Writing—review and editing: YM, HY, KK, KI, MT, TY, HT, K Ogasawara

## Conflicts of Interest

KK has received speaker honoraria from Daiichi Sankyo, Biotronik Japan, Pfizer, and Medtronic Japan and research grants (unrelated to this study) from Medtronic Japan and Biotronik Japan. HT has received honoraria for lectures and speaker bureaus from Daiichi Sankyo, Novartis, Medtronic Japan, Biotronik Japan, Bristol Myers Squibb, and Boston Scientific Japan and grants (investigator-initiated research unrelated to this study) from Abbott Laboratories Japan, Daiichi Sankyo, Nippon Boehringer Ingelheim, Otsuka Pharmaceutical, Eli Lilly Japan, and Marubun Tsusyo. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Goodness-of-fit statistics (Akaike information criterion and Bayesian information criterion) for candidate parametric survival models used for survival extrapolation, observed Kaplan-Meier curve and fitted parametric survival models, and coding definitions used to identify interventions and complications in the claims database.
[DOCX File, 121 KB - i-jmr_v15i1e94073_app1.docx ]

## References

1. Al-Majed NS, McAlister FA, Bakal JA, Ezekowitz JA. Meta-analysis: cardiac resynchronization therapy for patients with less symptomatic heart failure. Ann Intern Med 2011 Mar 15;154(6):401-412. [doi: 10.7326/0003-4819-154-6-201103150-00313] [Medline: 21320922]
2. Sudesh S, Abraham WT, Cleland JG, et al. Cardiac resynchronization therapy in ischemic versus nonischemic cardiomyopathy: patient-level meta-analysis of 7 randomized clinical trials. JACC Heart Fail 2024 Nov;12(11):1915-1924. [doi: 10.1016/j.jchf.2024.08.010] [Medline: 39387768]
3. Woods B, Hawkins N, Mealing S, et al. Individual patient data network meta-analysis of mortality effects of implantable cardiac devices. Heart 2015 Nov;101(22):1800-1806. [doi: 10.1136/heartjnl-2015-307634] [Medline: 26269413]
4. Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GM, Coats AJ. Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res 2023 Jan 18;118(17):3272-3287. [doi: 10.1093/cvr/cvac013] [Medline: 35150240]
5. Yokoshiki H, Shimizu A, Mitsuhashi T, et al. Trends in the use of implantable cardioverter-defibrillator and cardiac resynchronization therapy device in advancing age: analysis of the Japan cardiac device treatment registry database. J Arrhythm 2020 Jun;36(4):737-745. [doi: 10.1002/joa3.12377] [Medline: 32782648]
6. Okura Y, Ramadan MM, Ohno Y, et al. Impending epidemic: future projection of heart failure in Japan to the year 2055. Circ J 2008 Mar;72(3):489-491. [doi: 10.1253/circj.72.489] [Medline: 18296852]
7. Tsutsui H, Isobe M, Ito H, et al. JCS 2017/JHFS 2017 guideline on diagnosis and treatment of acute and chronic heart failure - digest version. Circ J 2019 Sep 25;83(10):2084-2184. [doi: 10.1253/circj.CJ-19-0342] [Medline: 31511439]
8. Zhang L, Ono Y, Qiao Q, Nagai T. Trends in heart failure prevalence in Japan 2014-2019: a report from healthcare administration databases. ESC Heart Fail 2023 Jun;10(3):1996-2009. [doi: 10.1002/ehf2.14321] [Medline: 37016908]
9. Tripoliti EE, Papadopoulos TG, Karanasiou GS, Naka KK, Fotiadis DI. Heart failure: diagnosis, severity estimation and prediction of adverse events through machine learning techniques. Comput Struct Biotechnol J 2017;15:26-47. [doi: 10.1016/j.csbj.2016.11.001] [Medline: 27942354]



XSL•FO
RenderX

10. Hadwiger M, Schumann L, Eisemann N, et al. A long-term cost-effectiveness analysis of cardiac resynchronisation therapy with or without defibrillator based on health claims data. Cost Eff Resour Alloc 2022 Sep 2;20(1):48. [doi: 10.1186/s12962-022-00384-x] [Medline: 36056371]
11. Mealing S, Woods B, Hawkins N, et al. Cost-effectiveness of implantable cardiac devices in patients with systolic heart failure. Heart 2016 Nov 1;102(21):1742-1749. [doi: 10.1136/heartjnl-2015-308883] [Medline: 27411837]
12. Shah D, Lu X, Paly VF, Tsintzos SI, May DM. Cost-effectiveness analysis of implantable cardiac devices in patients with systolic heart failure: a US perspective using real world data. J Med Econ 2020 Jul;23(7):690-697. [doi: 10.1080/13696998.2020.1746316] [Medline: 32207659]
13. Noda T, Ueda N, Tanaka Y, et al. Cost-effectiveness analysis of cardiac implantable electronic devices with reactive atrial-based antitachycardia pacing. Europace 2023 Mar 30;25(3):1087-1099. [doi: 10.1093/europace/euad003] [Medline: 36691793]
14. Noda T, Lu X, Ishiguro Y, et al. Cost-effective analysis of automated programming optimization in cardiac resynchronization therapy: Holistic Markov modelling. J Cardiol 2022 Jun;79(6):734-739. [doi: 10.1016/j.jjcc.2021.12.016] [Medline: 35016809]
15. Guideline for preparing cost-effectiveness evaluation to the Central Social Insurance Medical Council. National Institute of Public Health. 2024. URL: https://c2h.niph.go.jp/tools/guideline/guideline_en_2024.pdf [accessed 2025-07-22]
16. Sapp JL, Sivakumaran S, Redpath CJ, et al. Long-term outcomes of resynchronization-defibrillation for heart failure. N Engl J Med 2024 Jan 18;390(3):212-220. [doi: 10.1056/NEJMoa2304542] [Medline: 38231622]
17. Guyot P, Ades AE, Ouwens M, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol 2012 Feb 1;12:9. [doi: 10.1186/1471-2288-12-9] [Medline: 22297116]
18. Tang AS, Wells GA, Talajic M, et al. Cardiac-resynchronization therapy for mild-to-moderate heart failure. N Engl J Med 2010 Dec 16;363(25):2385-2395. [doi: 10.1056/NEJMoa1009540] [Medline: 21073365]
19. Tseng ZH, Hayward RM, Clark NM, et al. Sudden death in patients with cardiac implantable electronic devices. JAMA Intern Med 2015 Aug;175(8):1342-1350. [doi: 10.1001/jamainternmed.2015.2641] [Medline: 26098676]
20. Moss AJ, Schuger C, Beck CA, et al. Reduction in inappropriate therapy and mortality through ICD programming. N Engl J Med 2012 Dec 13;367(24):2275-2283. [doi: 10.1056/NEJMoa1211107] [Medline: 23131066]
21. Yokoshiki H, Watanabe M, Hamaguchi S, et al. Survival benefit of primary prevention implantable cardioverter-defibrillator/cardiac resynchronization therapy with a defibrillator: analysis of the Japan cardiac device treatment registry and Japanese cardiac registry of heart failure in cardiology. J Arrhythm 2025;41(3):e70084. [doi: 10.1002/joa3.70084] [Medline: 40357355]
22. Landolina M, Gasparini M, Lunati M, et al. Long-term complications related to biventricular defibrillator implantation: rate of surgical revisions and impact on survival: insights from the Italian Clinical Service Database. Circulation 2011 Jun 7;123(22):2526-2535. [doi: 10.1161/CIRCULATIONAHA.110.015024] [Medline: 21576653]
23. Køber L, Thune JJ, Nielsen JC, et al. Defibrillator implantation in patients with nonischemic systolic heart failure. N Engl J Med 2016 Sep 29;375(13):1221-1230. [doi: 10.1056/NEJMoa1608029] [Medline: 27571011]
24. Giacopelli D, Azzolina D, Comoretto RI, et al. Implantable cardioverter defibrillator lead performance: a systematic review and individual patient data meta-analysis. Int J Cardiol 2023 Feb 15;373:57-63. [doi: 10.1016/j.ijcard.2022.11.048] [Medline: 36460209]
25. Hayes D, Ramharter C, Paule S, Giacopelli D, Salehi B, Lang V. Worldwide data analysis of BIOTRONIK ICD and CRT-D battery longevity. J Interv Card Electrophysiol 2026 Jan;69(1):109-116. [doi: 10.1007/s10840-025-02141-5] [Medline: 41051440]
26. Regarding explanatory materials for the 2024 medical fee revision. Ministry of Health, Labour and Welfare Japan. 2024. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000196352_00012.html [accessed 2025-07-22]
27. Yasuda S, Nakao K, Nishimura K, et al. The current status of cardiovascular medicine in Japan - analysis of a large number of health records from a nationwide claim-based database, JROAD-DPC. Circ J 2016 Oct 25;80(11):2327-2335. [doi: 10.1253/circj.CJ-16-0196] [Medline: 27725417]
28. Kono Y, Katano S, Otaka Y, et al. Association of in-hospital cardiac rehabilitation on hospital-associated disability for octogenarian patients with acute myocardial infarction - an insight from the JROAD-DPC database. Circ Rep 2024;7(1):25-30. [doi: 10.1253/circrep.CR-24-0130] [Medline: 39802126]
29. [C2H1906] Summary of cost-effectiveness evaluation of Ivabradine (Coralan). Center for Outcomes Research and Economic Evaluation for Health, National Institute of Public Health. 2021. URL: https://c2h.niph.go.jp/results/C2H1906/C2H1906_Summary_Eng.pdf [accessed 2025-07-22]
30. Bristow MR, Saxon LA, Boehmer J, et al. Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. N Engl J Med 2004 May 20;350(21):2140-2150. [doi: 10.1056/NEJMoa032423] [Medline: 15152059]
31. Implantable cardioverter defibrillators and cardiac resynchronisation therapy for arrhythmias and heart failure. National Institute for Health and Care Excellence. 2014. URL: https://www.nice.org.uk/guidance/ta314 [accessed 2025-07-22]




XSL•FO
RenderX

32. Hadwiger M, Frielitz FS, Eisemann N, et al. Cardiac resynchronisation therapy in patients with moderate to severe heart failure in Germany: a cost-utility analysis of the additional defibrillator. Appl Health Econ Health Policy 2021 Jan;19(1):57-68. [doi: 10.1007/s40258-020-00571-y] [Medline: 32215877]
33. Fudim M, Dalgaard F, Friedman DJ, et al. Comorbidities and clinical response to cardiac resynchronization therapy: patient-level meta-analysis from eight clinical trials. Eur J Heart Fail 2024 Apr;26(4):1039-1046. [doi: 10.1002/ejhf.3029] [Medline: 37671601]
34. McMurray JJ, Packer M, Desai AS, et al. Angiotensin-neprilysin inhibition versus enalapril in heart failure. N Engl J Med 2014 Sep 11;371(11):993-1004. [doi: 10.1056/NEJMoa1409077] [Medline: 25176015]
35. McMurray JJ, Solomon SD, Inzucchi SE, et al. Dapagliflozin in patients with heart failure and reduced ejection fraction. N Engl J Med 2019 Nov 21;381(21):1995-2008. [doi: 10.1056/NEJMoa1911303] [Medline: 31535829]
36. Kanaoka K, Iwanaga Y, Takegawa K, et al. Current practice of guideline-directed medical therapy after acute heart failure hospitalization: a nationwide observational study. JACC Asia 2025 Jun;5(6):786-795. [doi: 10.1016/j.jacasi.2025.01.015] [Medline: 40473329]
37. Gordois A, Tanaka Y, Uenishi T, Yamaguchi H, Shoji A, Hill M. Cost-effectiveness analyses to assess the value of reactive atrial-based anti-tachycardia pacing for patients with pacemakers and defibrillators: an Australian private healthcare system perspective. J Arrhythm 2025 Apr;41(2):e70043. [doi: 10.1002/joa3.70043] [Medline: 40271387]
38. Yokoshiki H, Shimizu A, Mitsuhashi T, et al. Improved outcomes of cardiac resynchronization therapy with a defibrillator in systolic heart failure: analysis of the Japan cardiac device treatment registry database. J Arrhythm 2023;40(1):30-37. [doi: 10.1002/joa3.12952] [Medline: 38333398]


## Abbreviations

**AIC:** Akaike information criterion
**BIC:** Bayesian information criterion
**CA:** cardiac arrest
**COMPANION:** Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure
**CRT-D:** cardiac resynchronization therapy with defibrillator
**DSA:** deterministic sensitivity analysis
**HF:** heart failure
**HFrEF:** heart failure and reduced ejection fraction
**HR:** hazard ratio
**ICER:** incremental cost-effectiveness ratio
**JROAD-DPC:** Japanese Registry Of All cardiac and vascular Disease-Diagnostic Procedure Combination
**LBBB:** left bundle branch block
**MADIT-RIT:** Multicenter Automatic Defibrillator Implantation Trial–Reduce Inappropriate Therapy
**NYHA:** New York Heart Association
**OMT:** optimal medical therapy
**PSA:** probabilistic sensitivity analysis
**QALY:** quality-adjusted life year
**RAFT:** Resynchronization-Defibrillation for Ambulatory Heart Failure Trial
**rATP:** reactive atrial-based antitachycardia pacing
**SGLT2:** sodium-glucose cotransporter 2


*Edited by M Balcarras; submitted 26.Feb.2026; peer-reviewed by CP Lau, T Noda; revised version received 08.Jun.2026; accepted 15.Jun.2026; published 21.Jul.2026.*

*Please cite as:*
*Ohashi K, Watanabe M, Morii Y, Yokoshiki H, Kusano K, Imai K, Takagi M, Yamane T, Tada H, Ogasawara K*
*Cardiac Resynchronization Therapy With Defibrillator Using the JROAD-DPC Database: Cost-Effectiveness Analysis*
*Interact J Med Res 2026;15:e94073*
*URL: https://www.i-jmr.org/2026/1/e94073*
*doi:10.2196/94073*


© Kazuki Ohashi, Masaya Watanabe, Yasuhiro Morii, Hisashi Yokoshiki, Kengo Kusano, Katsuhiko Imai, Masahiko Takagi, Teiichi Yamane, Hiroshi Tada, Katsuhiko Ogasawara. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 21.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction



XSL•FO
RenderX

in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Comparative Analysis of Cardiometabolic Risk Profiles Among Hypertensive and Normotensive Adults: Cross-Sectional Study

Lebogang Faith Thaga[1*], MSc; Gudani Goodman Mukoma[1*], PhD; Terry Jeremy Ellapen[2], PhD; Takalani Clearance Muluvhu[2*], PhD

[1]Biokinetics, Recreation and Sport Science, Health Science, University of Venda, Thohoyandou, Limpopo, South Africa

[2]Department of Human Movement and Therapeutic Sciences, Faculty of Science, Tshwane University of Technology, Pretoria, Gauteng, South Africa

[*]these authors contributed equally

**Corresponding Author:**
Lebogang Faith Thaga, MSc
Biokinetics, Recreation and Sport Science
Health Science
University of Venda
University Road
Thohoyandou, Limpopo, 0950
South Africa
Phone: 27 719582990
Fax: 27 159629206
Email: lebogang.thaga@univen.ac.za

## Abstract

**Background:** Hypertension is a significant modifiable risk factor for cardiometabolic syndrome, leading to increased morbidity and premature mortality. Understanding the differences in cardiometabolic risk profiles between hypertensive and normotensive adults is essential for effective prevention strategies.

**Objective:** This study examined cardiometabolic risk profiles among adults in Johannesburg South, Gauteng Province, South Africa.

**Methods:** A cross-sectional study was conducted with 209 adults aged 25 to 65 years who were recruited from 3 Johannesburg townships in Region G. Anthropometric measures (BMI and waist circumference), blood pressure, and biochemical indexes (random glucose, total cholesterol, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, and triglycerides) were assessed. Statistical analyses were performed using Stata (version 19), with statistical significance set at $P<.05$.

**Results:** Hypertension prevalence increased significantly with BMI ($P<.001$) and waist circumference ($P=.003$). In multivariable analysis, age and elevated glucose levels were independently associated with hypertension, whereas obesity and central adiposity were not significantly associated after adjustment.

**Conclusions:** Hypertension was independently associated with advancing age and elevated glucose levels in this population. Although obesity and central adiposity showed significant associations in unadjusted analyses, they were not independently associated with hypertension after adjustment. These findings highlight the importance of early screening and targeted interventions to address cardiometabolic risk in underserved South African communities.

***(Interact J Med Res 2026;15:e84466)*** doi:10.2196/84466

**KEYWORDS**

adults; cardiometabolic risk; comparative analysis; hypertensive; South Africa; obesity

## Introduction

Hypertension remains a persistent global health challenge, significantly elevating the risk of cardiovascular and other noncommunicable diseases (NCDs). These conditions often lead to long-term disability. In 2017, NCDs accounted for 67% of all-age disability-adjusted life years in sub-Saharan Africa [1]. Commonly referred to as a silent killer, hypertension typically presents no obvious symptoms, making early detection difficult [2]. Regular screening is a reliable diagnostic method; however, it remains largely inaccessible in many low- and middle-income countries, including South Africa. As a result,



XSL•FO
RenderX

the prevalence of hypertension continues to rise, particularly in countries in the World Health Organization (WHO) African region, where it stands at 27%, notably higher than the 18% reported in the WHO American region [3].

With the rising prevalence of hypertension, research has increasingly focused on cardiometabolic risks, highlighting obesity as the most prominent predictor [4,5]. Onagbiye et al [6] identified abdominal obesity as a major contributor to cardiometabolic disease among university staff employees. Similarly, Modjadji et al [4] found that the odds of developing hypertension were high for obese truck drivers based on waist circumference (WC; adjusted odds ratio [AOR] 4.68, CI 1.92-11.34) and waist-to-height ratio (AOR 5.49, CI 1.74-17.27). Interestingly, a study among a younger population also concurred with these findings [7]. Despite consistent evidence across different populations, regions, and demographic groups, the global prevalence of overweight and obesity continues to increase.

It is important to note that a decade ago, central obesity was proven to be the key determinant of the prevalence of metabolic syndrome in sub-Saharan Africa [8]. Efforts to address this burden have encountered several challenges. For instance, a recent study in Kenya revealed a high prevalence of central obesity, with women being significantly more affected than men (58.3% compared to 35.6%) [9]. Furthermore, the 2025 WHO report highlights a continuing upward trend in global obesity rates, noting that 43% of adults aged ≥18 years are overweight and 16% are living with obesity. This growing trend is concerning, as obesity significantly increases the risk of developing type 2 diabetes mellitus, hyperlipidemia, hypertension, and other cardiovascular diseases [10]. These illnesses often coexist, increasing health risks and reducing life expectancy. Their combined effects extend beyond individuals; they have a substantial negative impact on the national economy, health care systems, and workplace productivity.

The increased cost of treating associated comorbid diseases related to obesity imposes a substantial financial burden on the South African health care system. In particular, overweight and obesity accounted for ZAR 28,734 million (ZAR 1=US $0.060 as of July 29, 2026) in expenses related to cardiovascular and endocrine diseases, with approximately 91% of this cost coming from hypertension and diabetes [11]. These figures are alarming, as access to quality health care varies widely across South Africa. Currently, 44% of all health care expenditure occurs in the private sector, serving only 16% of the South African population [12]. This means the rest of the population relies on a public health system that is underresourced and often inaccessible due to geographical, financial, or knowledge barriers [13].

Given this background, comprehensive profiling of cardiometabolic risk factors in underresourced communities is essential to inform targeted prevention and early intervention strategies. Although previous South African studies have demonstrated associations between obesity, central adiposity, and hypertension, most have focused on occupational cohorts, such as truck drivers and university staff, with limited evidence from community-dwelling adults living in underresourced urban settings. Consequently, there remains a significant gap in localized epidemiological data describing cardiometabolic risk profiles among township populations. This study addresses this gap by examining the cardiometabolic risk profiles of adults residing in selected townships in Johannesburg South. Specifically, the study aimed to compare the cardiometabolic risk profiles of hypertensive and normotensive adults and to identify independent predictors of hypertension.

## Methods

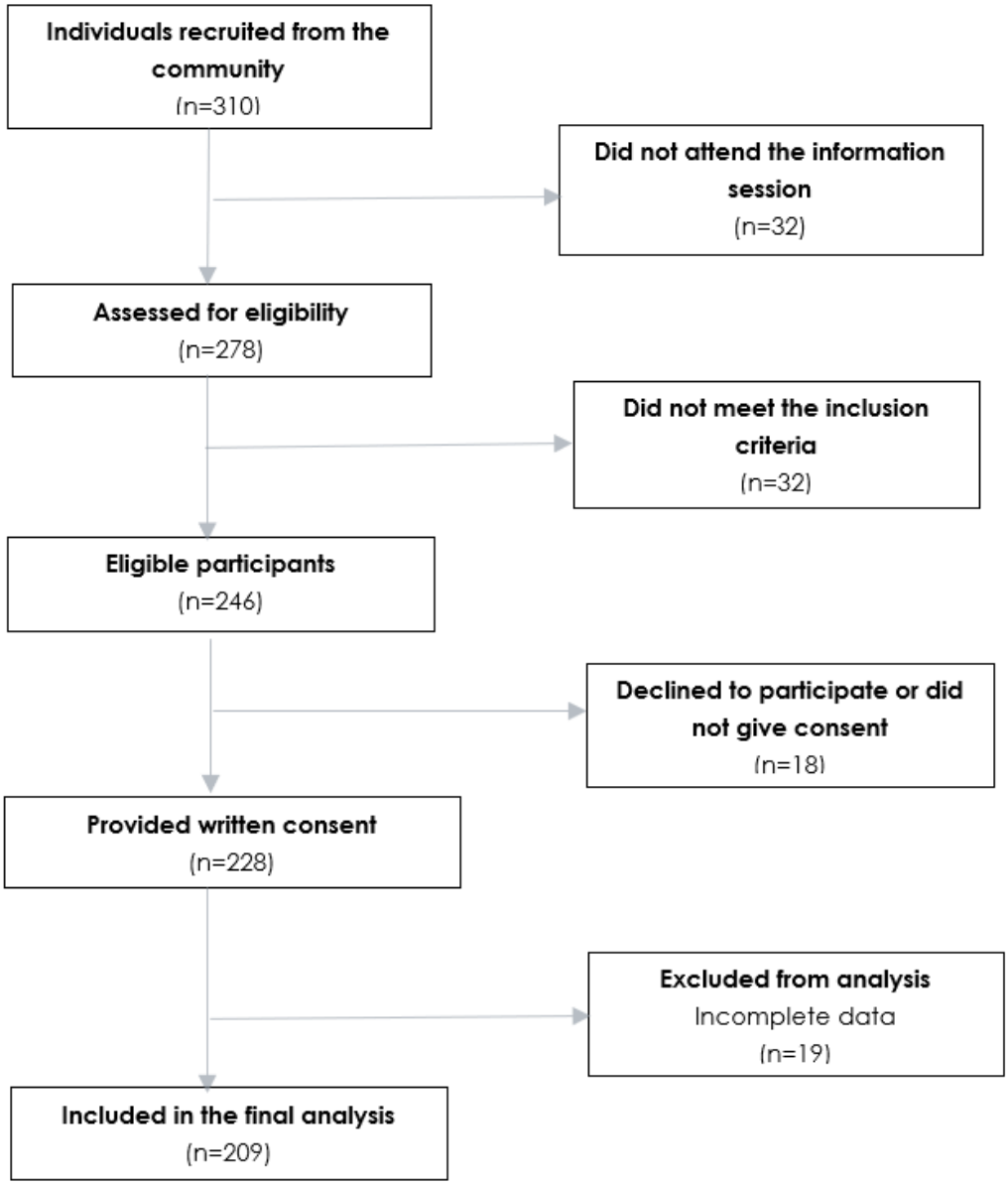
### Study Design and Participants

This cross-sectional study used purposive sampling to recruit male and female adults aged 25 to 65 years from selected townships in Johannesburg South. Although purposive sampling may limit representativeness and external validity, it was the most feasible approach because of the absence of reliable sampling frames in informal township settings. This approach enabled the inclusion of underrepresented community members and provided valuable baseline evidence to inform future probabilistic studies. A total of 310 individuals were recruited. Of these, 101 (32.6%) were excluded because they did not attend the information session (n=32, 10.3%), did not meet the inclusion criteria (n=32, 10.3%), declined to provide informed consent (n=18, 5.8%), or had incomplete anthropometric or biochemical data (n=19, 6.1%). The final analytic sample comprised 209 participants (n=72, 34.4% men and n=137, 65.6% women) with complete data for all study variables, including demographic characteristics, blood pressure, anthropometric measurements, and biochemical indexes. Missing data were handled using listwise deletion; therefore, no imputation methods were applied. Analyses were restricted to complete cases to maintain the consistency and reliability of statistical comparisons. Participant recruitment, eligibility assessment, enrollment, exclusion, and inclusion in the final analysis are summarized in Figure 1.



XSL•FO
RenderX

**Figure 1.** Flow diagram of participant recruitment, eligibility assessment, enrollment, exclusion, and inclusion in the final analysis.




Between 2023 and 2025, participants were recruited from community churches, households, and social gatherings in selected townships within Region G of the city of Johannesburg using a purposive sampling approach. Region G is the second most populated administrative region in the municipality and experiences considerable socioeconomic deprivation [14]. Eligible participants were permanent residents of the 3 selected townships and were aged 25 to 65 years.

The city of Johannesburg covers approximately 1645 $km^2$ in Gauteng Province. The central and northern suburbs are predominantly occupied by middle- and upper-income populations, whereas the southern suburbs and the far northern fringes are characterized by greater socioeconomic disadvantage. Approximately 20% of the city of Johannesburg population lives in informal settlements with limited access to basic municipal services, while a further 40% resides in inadequate housing with insufficient municipal support [15].

## Ethical Considerations

This study forms part of a PhD project approved by the Faculty Committee for Research Ethics–Science (FCRE-SCI), Tshwane University of Technology (FCRE-SCI 2023/05/016 [SCI; FCPS 02]). The study was conducted in accordance with the ethical principles of the Declaration of Helsinki and its subsequent amendments. Written informed consent was obtained from all participants prior to enrollment. Participants with elevated blood pressure, abnormal blood glucose levels, or other clinically significant findings identified during the screening were informed of their results and verbally advised to visit their nearest primary health care facility for further clinical evaluation and management.



XSL•FO
RenderX

### Instrumentation

#### *Anthropometric Measurements*

Anthropometric measurements were taken after participants completed the questionnaires, with assistance from research assistants or the principal investigator. Body mass was recorded to the nearest 0.1 kg, using a calibrated BFW 300 Platform Scale (Adam Equipment Co Ltd), and body stature to the nearest 0.1 cm, using the portable Seca 213 stadiometer (CE 0123) according to the validated protocols of the International Society for the Advancement of Kinanthropometry [16]. BMI was categorized according to the American College of Sports Medicine cut points [17] as follows: BMI<18.5 kg/$m^2$=underweight, BMI 18.5-24.9 kg/$m^2$=normal, BMI 25.0-29.9 kg/$m^2$=overweight, and BMI≥30.0 kg/$m^2$=obese.

WC measurements were recorded to the nearest 0.1 cm using a nonstretchable standard Lufkin tape measure manufactured by Cooper Tools of Apex. The WC risk measurements were categorized according to low risk (WC<80 cm in women and <94 cm in men) and high risk (WC≥88 cm in women and ≥94 cm in men) [18].

#### *Biochemical Measurements*

Blood pressure was measured with the arm supported at the heart level (European Society of Hypertension) with a sphygmomanometer (Omron) using the Riva-Rocci or Korotkoff method on the left arm [19]. An appropriately sized blood pressure cuff was used, and participants were instructed to refrain from smoking cigarettes and from ingesting caffeine for 30 minutes prior to measurements. Measurements were taken thrice with a 5-minute resting period between each measurement [20]. The average of the measurements was used for data analysis.

The blood pressure cutoff points used for analysis were assessed according to guidelines of the American College of Cardiology and American Heart Association, where normal blood pressure is defined as <120 mm Hg systolic and <80 mm Hg diastolic, elevated blood pressure as 120 to 129 mm Hg systolic and <80 mm Hg diastolic, stage 1 hypertension as 130 to 139 mm Hg systolic or 80 to 89 mm Hg diastolic, and stage 2 hypertension as ≥140 mm Hg systolic or ≥90 mm Hg diastolic [19]. Participants who were on hypertensive medication with a known diagnosis were considered hypertensive. Participants who presented with hypertension at baseline and who were not on medication were referred to the clinic.

A peripheral blood sample was collected to determine random glucose levels, total cholesterol, high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and triglycerides. The values were recorded in millimoles per liter according to the manufacturer's CardioChek Plus analyzer user guide v1.12 or later. Random glucose levels <7.8 mmol/L were classified as low risk and ≥7.8 mmol/L as high risk. Random glucose was a practical measure in township settings where fasting requirements and laboratory access for hemoglobin $A_{1c}$ ($HbA_{1c}$) testing were not feasible, whereas less precise, random glucose testing would have allowed immediate, point-of-care screening and referral, thereby ensuring participation from underresourced communities. Cholesterol norms were classified as total cholesterol<5 mmol/L=normal, LDL-C<3 mmol/L=normal, HDL-C<1.04-mmol/L=low, and HDL-C≥1.2 mmol/L=optimal. Triglyceride concentrations were classified as normal (<1.7 mmol/L) and high risk (≥1.7 mmol/L) [18]. Persons taking antiglycemic medication were also classified as diabetic.

### Statistical Analysis

Data were cleaned, coded, and analyzed using Stata SE (version 19; StataCorp). The distribution of continuous variables was assessed using the Shapiro-Wilk test, histograms, and Q-Q plots to evaluate normality and guide the selection of appropriate descriptive statistics. As the continuous variables were not normally distributed, they are presented as medians with IQR. Categorical variables are presented as frequencies and percentages. The prevalence of cardiometabolic risk factors and hypertension was summarized using proportions.

A multivariable logistic regression (LR) model was used to identify independent predictors of hypertension. Prior to model fitting, the assumptions of LR were evaluated. The outcome variable was binary (hypertensive vs normotensive), observations were independent, multicollinearity was assessed using variance inflation factors, and model fit was evaluated using the Hosmer-Lemeshow goodness-of-fit test. Variance inflation factor values ranged from 1.11 to 1.93, indicating no evidence of problematic multicollinearity among the predictors included in the model. Statistical significance was set at $P$<.05.

## Results

### Descriptive Statistics

The baseline characteristics of the 209 study participants are summarized in Table 1. The median age was 45 (IQR 32-57) years. Additional anthropometric and clinical characteristics included a median BMI of 27.6 (IQR 15.8-52.7) kg/$m^2$ and a median systolic blood pressure (SBP) of 122 (IQR 65-181) mm Hg.



XSL•FO
RenderX

**Table 1.** Demographic and anthropometric characteristics (N=209).

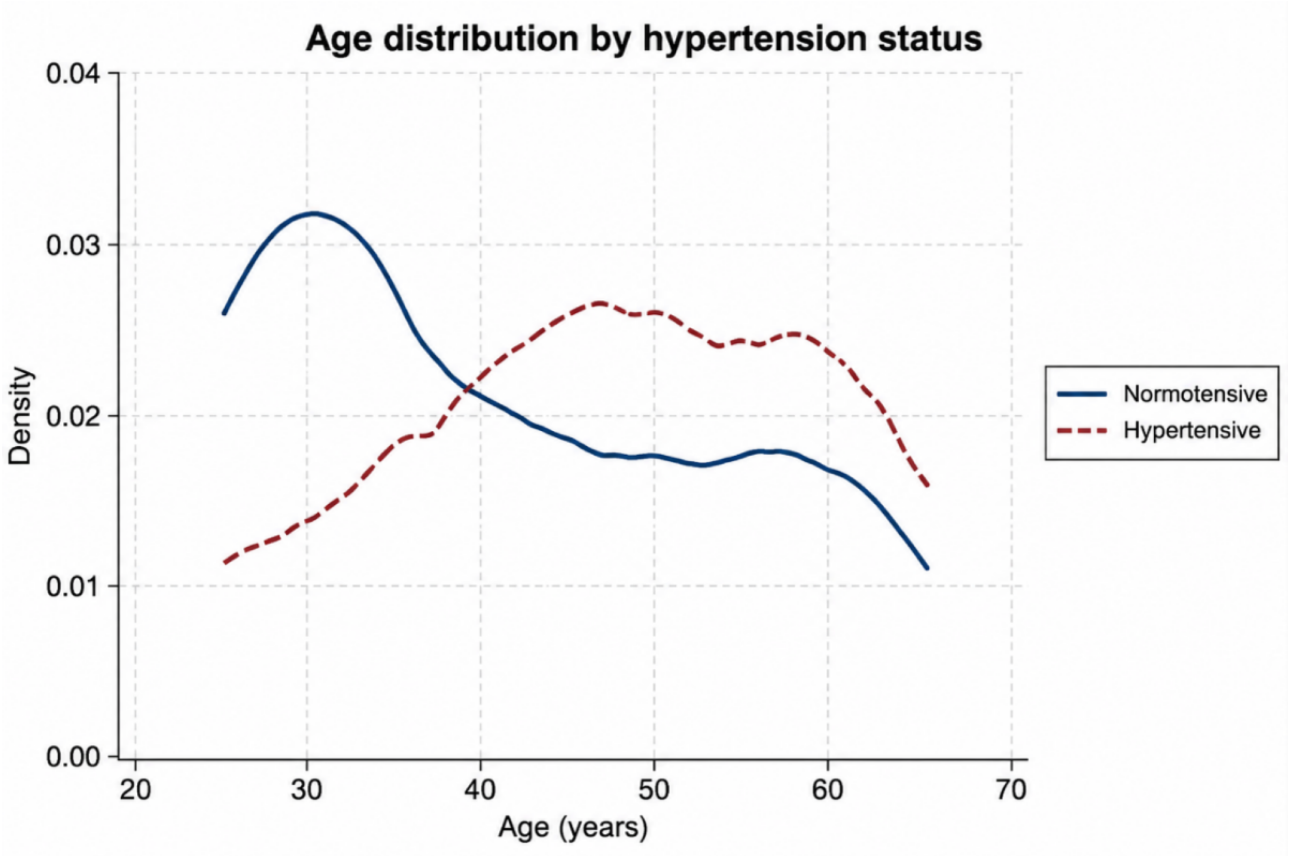
| Variables | Median (IQR) |
|---|---|
| Age (years) | 45 (32-57) |
| BMI (kg/m$^2$) | 27.6 (24.1-33) |
| Waist circumference (cm) | 94 (82-106) |
| Systolic blood pressure (mm Hg) | 122 (112-134) |
| Diastolic blood pressure (mm Hg) | 80 (71-89) |
| Glucose (mmol/L) | 5.8 (5.2-7) |
| Total cholesterol (mmol/L) | 3.76 (3.18-4.46) |
| HDL[a] cholesterol (mmol/L) | 1.11 (0.91-1.37) |
| LDL[b] cholesterol (mmol/L) | 1.97 (1.11-2.48) |
| Triglycerides (mmol/L) | 1.22 (0.8-1.76) |

[a]HDL: high-density lipoprotein.

[b]LDL: low-density lipoprotein.

Figure 2 presents the age distribution of participants according to hypertension status. The normotensive group, represented by the blue curve, peaked in the late 20s to early 30s and gradually declined with increasing age. In contrast, the hypertensive group (red curve) showed increasing density with advancing age, peaking in the late 50s to early 60s. The 2 distributions intersected at approximately 45 to 50 years, suggesting a shift in the age range at which hypertension became more prevalent than normotension. These findings support the well-established association between advancing age and increased hypertension prevalence.

**Figure 2.** Density plot showing the age distribution of normotensive and hypertensive participants. The blue curve represents normotensive participants, and the red curve represents hypertensive participants.






XSL•FO
RenderX

### Association Between Obesity, Central Obesity, and Hypertension Status

Table 2 presents the association between BMI, WC, and hypertension status, which was analyzed using chi-square tests. A statistically significant association was observed between BMI category and hypertension (*P*<.001). The prevalence of hypertension increased progressively with higher BMI, rising from 30.4% (17/56) among individuals with normal BMI to 64.6% (53/82) among those classified as obese. Similarly, WC showed a significant association with hypertension (*P*=.003), with 53.4% (79/148) of participants with high WC being hypertensive, compared to 31.1% (19/61) of those with normal WC.

**Table 2.** Association between obesity, central obesity, and hypertension status among study participants (N=209).

| Variable and category | Normotensive, n (%) | Hypertensive, n (%) | *P* value[a] |
|---|---|---|---|
| **BMI** | | | <.001 |
| Underweight | 4 (50) | 4 (50) | |
| Normal weight | 39 (69.6) | 17 (30.4) | |
| Overweight | 39 (61.9) | 24 (38.1) | |
| Obese | 29 (35.4) | 53 (64.6) | |
| **Waist circumference** | | | .003 |
| Normal | 42 (68.9) | 19 (31.1) | |
| High | 69 (46.6) | 79 (53.4) | |

[a]*P* values were calculated using Pearson chi-square test.

### Lipid Profiles

The association between several lipid categories and the presence of hypertension is presented in Table 3. Chi-square test results for these variables indicate no statistically significant association with hypertension. The *P* values for HDL cholesterol (*P*=.65), LDL cholesterol (*P*=.70), and triglycerides (*P*=.25) are well above the conventional significance level of .05. This means that, in this dataset, the distribution of hypertension across the categories for these lipid variables is not different from what would be expected by chance.

**Table 3.** Comparison of lipid profile categories between normotensive and hypertensive participants (N=209).

| Variable and category | Normotensive (n=111), n (%) | Hypertensive (n=98), n (%) | Chi-square (*df*) | *P* value[a] |
|---|---|---|---|---|
| **HDL[b] cholesterol** | | | 0.2 (1) | .65 |
| Normal | 43 (51.2) | 41 (48.8) | | |
| High risk | 68 (54.4) | 57 (45.6) | | |
| **LDL[c] cholesterol** | | | 0.1 (1) | .70 |
| Normal | 99 (52.7) | 89 (47.3) | | |
| High risk | 12 (57.1) | 9 (42.9) | | |
| **Triglycerides** | | | 1.3 (1) | .25 |
| Normal | 87 (55.4) | 70 (44.6) | | |
| High risk | 24 (46.2) | 28 (53.8) | | |
| **Total cholesterol** | | | 3.3 (1) | .07 |
| Normal | 102 (55.4) | 82 (44.6) | | |
| High risk | 9 (36) | 16 (64) | | |

[a]*P* values were calculated using Pearson chi-square test.

[b]HDL: high-density lipoprotein.

[c]LDL: low-density lipoprotein.

Interestingly, the results for total cholesterol show a borderline association with hypertension. While the *P* value of .07 is slightly above the .05 significance threshold, the association was not statistically significant. The table shows that a lower percentage of participants in the high-risk total cholesterol category (9/25, 36%) had no hypertension, and a higher rate (16/25, 64%) had hypertension, compared with participants in the normal total cholesterol category. This suggests that having a higher total cholesterol level might be associated with an increased likelihood of hypertension, but the finding is not statistically significant at the 5% level.



XSL•FO
RenderX

## Predictors of Hypertension

A multivariable LR model was fitted to identify predictors of hypertension status. The model was statistically significant (LR: $\chi^2_{11}$=76.3; *P*<.001), indicating that the predictors collectively explained variation in hypertension. The LR model demonstrated adequate fit to the data based on the Hosmer-Lemeshow goodness-of-fit test ($\chi^2_8$=7.2; *P*=.52).

As shown in Table 4, age and elevated glucose levels were independently associated with hypertension. Increasing age was positively associated with hypertension (β=0.07, 95% CI 0.041-0.099; *P*<.001). Elevated glucose levels were also independently associated with hypertension (β=2.277, 95% CI 0.684-3.871; *P*<.005). Other variables, including gender, BMI category, WC, and lipid levels, were not significantly associated with hypertension after adjustment. The crude associations observed for BMI and WC in the bivariate analyses may reflect their correlation with advancing age.

**Table 4.** Multivariable logistic regression (LR) analysis of independent predictors of hypertension.

| Variable | β (95% CI) | *P* value |
|---|---|---|
| **Individual-level factors** | | |
| Age (years) | 0.07 (0.041 to 0.099) | <.001 |
| Sex (female vs male) | 0.155 (−0.677 to 0.987) | .71 |
| **Metabolic factors** | | |
| **BMI category** | | |
| Normal weight | −1.51 (−3.207 to 0.186) | .08 |
| Overweight | −1.238 (−3.031 to 0.556) | .18 |
| Obese | −0.058 (−1.874 to 1.757) | .95 |
| Waist circumference (high vs normal) | 0.124 (−0.966 to 1.214) | .82 |
| HDL[b] cholesterol (high vs normal) | −0.219 (−0.986 to 0.549) | .58 |
| LDL[c] cholesterol (high vs normal) | −1.043 (−2.286 to 0.200) | .10 |
| Triglycerides (high vs normal) | −0.178 (−0.984 to 0.628) | .67 |
| Total cholesterol (high vs normal) | 0.806 (−0.518 to 2.130) | .23 |
| Glucose category (high vs normal) | 2.277 (0.684 to 3.871) | .005 |

[b]HDL: high-density lipoprotein.

[c]LDL: low-density lipoprotein.

Statistical significance was set at *P*<.05. Model fit: likelihood ratio $\chi^2_{11}$=76.3.

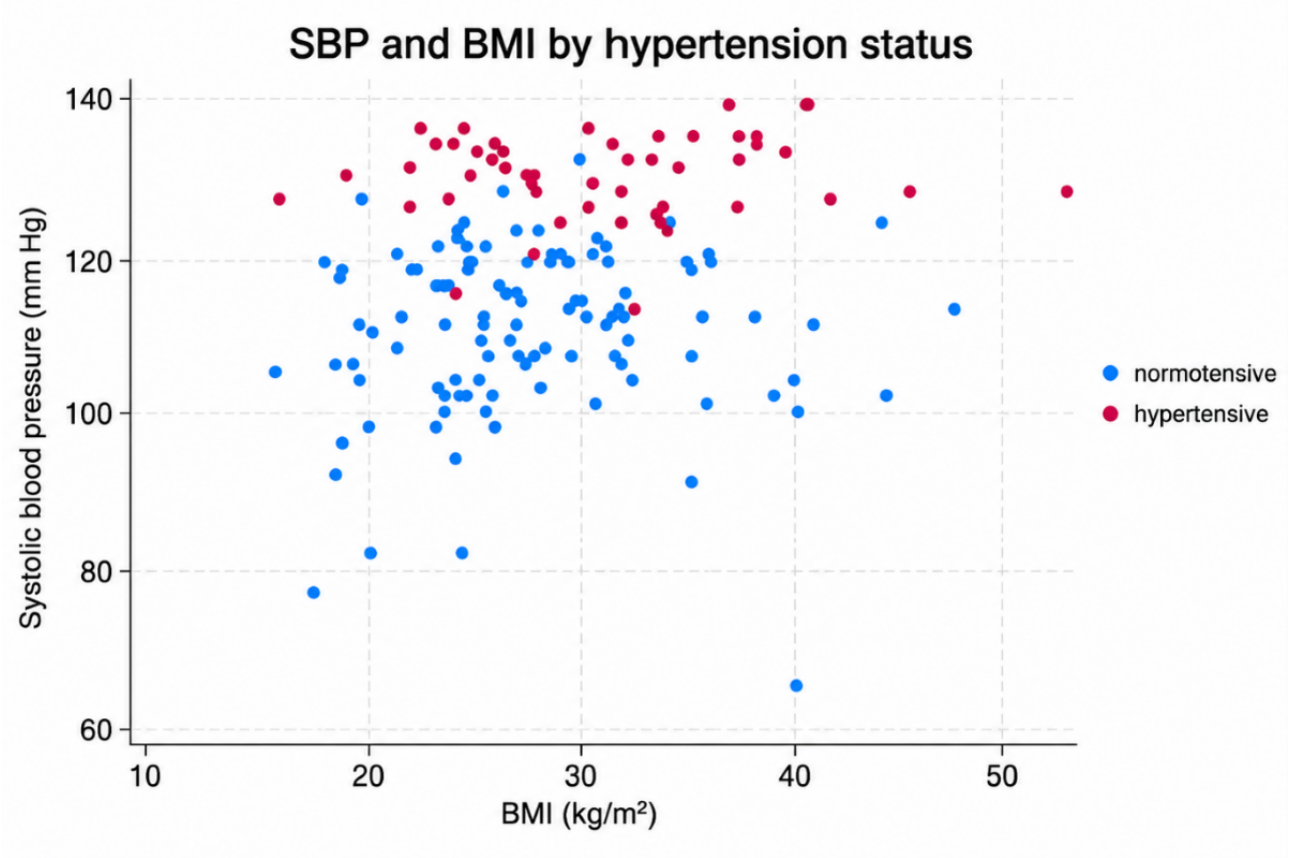
## Association Between SBP, BMI, and Hypertension

Figure 3 illustrates the relationship between SBP and BMI according to hypertension status. Hypertensive participants consistently exhibited higher SBP values than normotensive participants across the BMI spectrum. Although higher BMI values were observed in both groups, substantial overlap was evident, suggesting that BMI alone did not clearly distinguish hypertension status in this population.



XSL•FO
RenderX

**Figure 3.** Scatter plot showing the relationship between systolic blood pressure (SBP) and BMI among normotensive and hypertensive participants. Blue points represent normotensive participants, and red points represent hypertensive participants.




## Discussion

### Principal Findings

This study compared cardiometabolic risks between hypertensive and normotensive adults in selected townships of Johannesburg. The primary finding of this study was that older age and elevated glucose levels were the strongest independent predictors of hypertension in the adjusted models. Our analysis showed that increasing age and elevated glucose levels were both independently associated with hypertension. These findings align with established evidence linking advancing age to vascular stiffening and other physiological changes that contribute to elevated blood pressure [21,22].

Although obesity and central adiposity showed significant associations with hypertension in crude analyses, these associations were attenuated primarily in adjusted models. This suggests that the relationship between obesity and hypertension in this population is primarily mediated by age, which plays a dominant role. Consistent with longitudinal evidence, the risk of developing hypertension increases steadily with advancing age [23]. At the same time, cohort studies consistently demonstrate that elevated BMI and WC are strongly linked to hypertension risk across all stages of life, underscoring the critical role of excess adiposity in disease development [23,24].

Notably, life stage–specific research has shown that high baseline BMI is the strongest predictor of hypertension in early adulthood, whereas high WC exerts the greatest impact during childhood [25]. The high prevalence of obesity among hypertensive individuals in our study (53/82, 64.6%) is clinically significant and mirrors global trends [26-28]. While other studies emphasize obesity as a key independent risk factor [29-31], our findings highlight the need to account for the strong independent effect of age when designing targeted interventions for this community.

From a public health perspective, these results suggest that interventions should focus on 2 key areas: addressing excess body fat across the lifespan and tailoring strategies to the specific risk profiles of different age groups. For younger populations, prevention efforts should prioritize weight management, physical activity, and healthy dietary habits to help prevent the onset of hypertension. In contrast, for older adults, interventions should focus on regular blood pressure screening, management of existing health conditions, and lifestyle modifications that are practical and achievable in later life. Community-based programs that combine age-sensitive approaches with obesity prevention strategies could have the greatest impact.

In contrast to other studies that reported dyslipidemia as a prevalent comorbidity [32-34], we did not find statistically significant independent associations between lipid markers (HDL, LDL, and triglycerides) and hypertension. However, the *P* value for total cholesterol was close to the conventional significance threshold ($P$=.07), suggesting a potential trend warranting further investigation. Although these lipid parameters did not reach statistical significance, their prevalence in the population remains clinically relevant and should be considered in risk reduction efforts.



XSL•FO
RenderX

Evidence from larger cross-sectional studies has consistently demonstrated a strong relationship between hypertension and dyslipidemia. For example, a study conducted in Poland reported a significant association between hypertension and abnormal cholesterol levels ($P$<.001), with hypertensive individuals showing a 3-fold increase in the odds of having abnormal cholesterol levels [35]. Similar trends have been observed in sub-Saharan Africa, where dyslipidemia frequently coexists with hypertension and contributes to elevated cardiometabolic risk [33,36]. In light of these findings, the absence of statistically significant associations in our study may be due to sample size limitations, population-specific characteristics, or unmeasured confounding factors, such as diet, physical activity, and socioeconomic status.

The South African context magnifies the importance of these findings. Research from other local populations has demonstrated strong correlations between BMI, WC, and SBP [6,37], mirroring the current study's outcomes. International evidence further highlights the growing burden of obesity, particularly among middle-aged adults and rural populations [9,28,30]. Together, these results emphasize the urgency of community-level interventions that combine regular health screenings, obesity management, and culturally tailored health education.

Overall, the findings suggest a clustering of cardiometabolic risk factors, with obesity and glucose dysregulation playing central roles in hypertension pathogenesis. The substantial proportion of undiagnosed or uncontrolled hypertension observed in Johannesburg South underscores the need for early detection, integrated screening, and multifactorial interventions.

## Limitations

Several limitations must be acknowledged. First, the cross-sectional design precludes the determination of causal relationships. The observed relationships should be interpreted as correlations rather than causal pathways. Longitudinal studies are necessary to establish temporal and causal links between age, glucose dysregulation, obesity, and hypertension.

Second, although purposive sampling limits representativeness and external validity, it was considered the most feasible approach because of the absence of reliable sampling frames in township settings. This method ensured the inclusion of underrepresented community members, providing valuable baseline evidence to inform future probabilistic studies.

Third, the final sample of 209 participants may have limited statistical power to detect modest associations, particularly for lipid-related variables and within the multivariable LR model. Consequently, some nonsignificant findings may reflect insufficient statistical power rather than the absence of a true association. Larger studies are needed to confirm these findings and improve the precision of the estimates. Furthermore, excluding participants with incomplete data or those who did not return for screening may have introduced selection bias, potentially limiting the representativeness of the final analytic sample. In addition, the underrepresentation of men, which is common in South African epidemiological research, may limit the generalizability of the findings.

Additionally, glycemic status was assessed using random blood glucose measurements rather than fasting plasma glucose or $HbA_{1c}$. Although random glucose testing was the most feasible approach in this community-based setting, it is less precise than standard diagnostic measures and may have resulted in misclassification of glycemic status.

Finally, the study did not include socioeconomic and behavioral factors, such as dietary intake, physical activity, and stress, which may have influenced the observed associations.

## Recommendations

Future research should use longitudinal designs with larger, more diverse samples to strengthen causal inference and generalizability. Incorporating dietary, lifestyle, and socioeconomic data will provide a more comprehensive understanding of modifiable risk factors. Moreover, the use of sensitive lipid biomarkers is essential to elucidate their role in hypertension. Evaluating community-based screening and health education programs could further inform effective prevention strategies.

## Conclusions

Hypertension was independently associated with advancing age and elevated glucose levels in this population. Although obesity and central adiposity showed significant associations in unadjusted analyses, they were not independently associated with hypertension after adjustment. While lipid markers were not statistically significant predictors, their prevalence remains clinically relevant and warrants continued monitoring. These findings highlight the importance of early screening and multifactorial interventions to address cardiometabolic risk in underserved South African communities.

## Acknowledgments

The authors would like to thank all the participants who provided written consent to participate in this study. The authors would also like to thank Rifiloe Moropane and Aviwe Mantshuphu for their assistance with data collection. The authors used generative AI during manuscript preparation. In accordance with the Generative AI Delegation Taxonomy (2025), the authors declare that Microsoft Copilot was used under full human oversight solely for proofreading, language editing, and improving readability. The authors reviewed and verified all AI-assisted outputs and accept full responsibility for the accuracy, integrity, and content of the final manuscript. Generative AI tools are not listed as authors and bear no responsibility for the published work.



XSL•FO
RenderX

## Funding

The corresponding author received institutional support for their PhD studies, which facilitated data collection for this research. The study was funded through the University of Venda's University Capacity Development Grant (2024-2025) and the Tshwane University of Technology Postgraduate Scholarship Program (2022-2023). The funding bodies had no role in the design, analysis, or interpretation of the study, and the views, recommendations, and conclusions presented are solely those of the authors.

## Authors' Contributions

LFT conceptualized the study, designed the methodology, collected the data, and drafted the manuscript. GGM contributed to drafting the manuscript and was involved in reviewing, analyzing, and interpreting the results. TCM and TJE assisted in drafting and reviewing the manuscript.

## Conflicts of Interest

None declared.

## References

1. Gouda HN, Charlson F, Sorsdahl K, Ahmadzada S, Ferrari AJ, Erskine H, et al. Burden of non-communicable diseases in sub-Saharan Africa, 1990–2017: results from the Global Burden of Disease Study 2017. Lancet Glob Health 2019 Oct;7(10):e1375-e1387. [doi: 10.1016/s2214-109x(19)30374-2]
2. Global report on hypertension: the race against a silent killer. World Health Organization. 2023. URL: https://www.who.int/publications/i/item/9789240081062 [accessed 2026-08-05]
3. Hypertension. World Health Organization. 2025. URL: https://www.who.int/news-room/fact-sheets/detail/hypertension [accessed 2025-09-15]
4. Modjadji P, Bokaba M, Mokwena KE, Mudau TS, Monyeki KD, Mphekgwana PM. Obesity as a risk factor for hypertension and diabetes among truck drivers in a logistics company, South Africa. Appl Sci 2022 Feb 07;12(3):1685. [doi: 10.3390/app12031685]
5. Micklesfield LK, Kolkenbeck-Ruh A, Mukoma G, Prioreschi A, Said-Mohamed R, Ware LJ, et al. The Healthy Aging Adult South Africa report card: a systematic review of the evidence between 2013 and 2020 for middle-aged South African men and women. Cardiovasc J Afr 2022;33(4):200-219 [FREE Full text] [doi: 10.5830/CVJA-2022-015] [Medline: 35789240]
6. Onagbiye SO, Smithdorf G, Ghaleelullah A, Andrews B, Young M, Bassett SH, et al. Prevalence of selected risk factors for cardiometabolic disease among university staff in the Western Cape, South Africa. Open Public Health J 2021 Nov 17;14:509-516. [doi: 10.2174/1874944502114010509]
7. Myette RL, Flynn JT. The ongoing impact of obesity on childhood hypertension. Pediatr Nephrol 2024 Aug;39(8):2337-2346. [doi: 10.1007/s00467-023-06263-8] [Medline: 38189961]
8. Fezeu L, Balkau B, Kengne AP, Sobngwi E, Mbanya JC. Metabolic syndrome in a sub-Saharan African setting: central obesity may be the key determinant. Atherosclerosis 2007 Jul;193(1):70-76 [FREE Full text] [doi: 10.1016/j.atherosclerosis.2006.08.037] [Medline: 17011567]
9. Nyakundi C, Okemwa S, Ngesa RW, Gatimu SM. Prevalence and determinants of central obesity among adults 18–69 years in Kenya: a cross-sectional study. medRxiv. Preprint posted online on September 19, 2024 2026. [doi: 10.1101/2024.09.18.24313881]
10. Chung PC, Hu TH, Chiao CH, Hwang JS, Chan TC. The long-term effects of cardiometabolic risk factors on mortality and life expectancy: evidence from a health check-up cohort study. BMC Cardiovasc Disord 2025 Jan 16;25(1):27 [FREE Full text] [doi: 10.1186/s12872-025-04469-2] [Medline: 39819280]
11. Boachie MK, Thsehla E, Immurana M, Kohli-Lynch C, Hofman KJ. Estimating the healthcare cost of overweight and obesity in South Africa. Glob Health Action 2022 Dec 31;15(1):2045092 [FREE Full text] [doi: 10.1080/16549716.2022.2045092] [Medline: 35389331]
12. van Rensburg HC. South Africa's protracted struggle for equal distribution and equitable access - still not there. Hum Resour Health 2014 May 08;12:26 [FREE Full text] [doi: 10.1186/1478-4491-12-26] [Medline: 24885691]
13. Morris LD, Grimmer KA, Twizeyemariya A, Coetzee M, Leibbrandt DC, Louw QA. Health system challenges affecting rehabilitation services in South Africa. Disabil Rehabil 2021 Mar;43(6):877-883. [doi: 10.1080/09638288.2019.1641851] [Medline: 31378096]
14. Joburg demographics and key socio economic indicators. Joburg. 2020. URL: https://joburg.org.za/documents_/Documents/Statistical%20Briefs/Issue%2023%20Joburg%20Demographics%20and%20Key%20Socio%20Economic%20Indicators.pdf [accessed 2025-09-15]
15. Profile and analysis district development model. Department of Cooperative Governance and Traditional Affairs (CoGTA), South Africa. 2020. URL: https://www.cogta.gov.za/ddm/wp-content/uploads/2020/11/City-of-Johannesburg-October-2020 [accessed 2025-09-15]



XSL•FO
RenderX

16. Esparza Ros F, Vaquero-Cristóbal R, Marfell-Jones M, Marfell-Jones M. International Standards for Anthropometric Assessment. Murcia, Spain: International Society for the Advancement of Kinanthropometry; 2019.
17. Ozemek C, American College of Sports Medicine, Bonikowske A, Christle J, Gallo P. ACSM's Guidelines for Exercise Testing and Prescription. Waltham, MA: Wolters Kluwer Health; 2025.
18. Ozemek C, Bonikowske A, Christle J, Gallo P. ACSM's Guidelines for Exercise Testing and Prescription. Philadelphia, PA: Lippincott Williams & Wilkins; 2025.
19. Whelton PK, Carey RM. The 2017 clinical practice guideline for high blood pressure. JAMA 2017 Dec 05;318(21):2073-2074. [doi: 10.1001/jama.2017.18209] [Medline: 29159375]
20. WHO technical specifications for automated non-invasive blood pressure measuring devices with cuff. World Health Organization. 2020 Apr 20. URL: https://www.who.int/publications/i/item/9789240002654: [accessed 2025-09-15]
21. Boutouyrie P, Chowienczyk P, Humphrey JD, Mitchell GF. Arterial stiffness and cardiovascular risk in hypertension. Circ Res 2021 Apr 02;128(7):864-886 [FREE Full text] [doi: 10.1161/CIRCRESAHA.121.318061] [Medline: 33793325]
22. Mitchell GF. Arterial stiffness in aging: does it have a place in clinical practice? Recent advances in hypertension. Hypertension 2021 Mar 03;77(3):768-780. [doi: 10.1161/HYPERTENSIONAHA.120.14515] [Medline: 33517682]
23. Li Z, Cao L, Zhou Z, Han M, Fu C. Factors influencing the progression from prehypertension to hypertension among Chinese middle-aged and older adults: a 2-year longitudinal study. BMC Public Health 2023 Feb 15;23(1):339 [FREE Full text] [doi: 10.1186/s12889-022-14410-3] [Medline: 36793011]
24. Jia G, Sowers JR, Whaley-Connell AT. Obesity in hypertension: the role of the expanding waistline over the years and insights into the future. Hypertension 2024 Apr;81(4):687-690 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.123.21719] [Medline: 38018438]
25. Meng X, Chen L, Qi Q, Liu W, Xue Y, Yan Y. Life stage-specific heterogeneity in the effects of adiposity transition on hypertension incidence and resolution. Int J Obes (Lond) 2026 Jun;50(6):1283-1289. [doi: 10.1038/s41366-026-02064-9] [Medline: 41922507]
26. Kandala NB, Nnanatu CC, Dukhi N, Sewpaul R, Davids A, Reddy SP. Mapping the burden of hypertension in South Africa: a comparative analysis of the National 2012 SANHANES and the 2016 Demographic and Health Survey. Int J Environ Res Public Health 2021 May 19;18(10):5445 [FREE Full text] [doi: 10.3390/ijerph18105445] [Medline: 34069668]
27. Lauder L, Mahfoud F, Azizi M, Bhatt DL, Ewen S, Kario K, et al. Hypertension management in patients with cardiovascular comorbidities. Eur Heart J 2023 Jun 20;44(23):2066-2077 [FREE Full text] [doi: 10.1093/eurheartj/ehac395] [Medline: 36342266]
28. Tu J, Chen H, Zeng Q, Chen L, Guo Y, Chen K. Trends in obesity prevalence among adults with hypertension in the United States, 2001 to 2023. Hypertension 2025 Mar;82(3):498-508. [doi: 10.1161/HYPERTENSIONAHA.124.24123] [Medline: 39758027]
29. Sørensen TI, Martinez AR, Jørgensen TS. Epidemiology of obesity. In: Eckel J, Clément K, editors. From Obesity to Diabetes. Handbook of Experimental Pharmacology. Cham, Switzerland: Springer; 2022:3-27.
30. WHO European regional obesity report 2022. World Health Organization. 2022 May 02. URL: https://www.who.int/europe/publications/i/item/9789289057738: [accessed 2025-09-19]
31. Kibret KT, Strugnell C, Backholer K, Peeters A, Tegegne TK, Nichols M. Life-course trajectories of body mass index and cardiovascular disease risks and health outcomes in adulthood: systematic review and meta-analysis. Obes Rev 2024 Apr;25(4):e13695. [doi: 10.1111/obr.13695] [Medline: 38226403]
32. Katundu KG, Mukhula V, Phiri T, Phiri C, Filisa-Kaphamtengo F, Chipewa P, et al. High prevalence of dyslipidaemia among persons with diabetes mellitus and hypertension at a tertiary hospital in Blantyre, Malawi. BMC Cardiovasc Disord 2022 Dec 22;22(1):557 [FREE Full text] [doi: 10.1186/s12872-022-03011-y] [Medline: 36544081]
33. Masilela C, Adeniyi OV, Benjeddou M. Prevalence, patterns and determinants of dyslipidaemia among South African adults with comorbidities. Sci Rep 2022 Jan 10;12(1):337 [FREE Full text] [doi: 10.1038/s41598-021-04150-6] [Medline: 35013433]
34. Obsa MS, Ataro G, Awoke N, Jemal B, Tilahun T, Ayalew N, et al. Determinants of dyslipidemia in Africa: a systematic review and meta-analysis. Front Cardiovasc Med 2022 Feb 23;8:778891 [FREE Full text] [doi: 10.3389/fcvm.2021.778891] [Medline: 35284497]
35. Kacprzak P, Michalak B, Nawrot A, Kierzkiewicz M, Walędziak M, Różańska-Walędziak A. Prevalence of dyslipidemia, arterial hypertension, and other risk factors in patients after cardiovascular events — a retrospective study in a Polish population. Arter Hypertens 2026 Apr 16;30:e01826002. [doi: 10.5603/ah.109752]
36. Katundu K. BMC Cardiovascular Disorders 2022;22:557. [doi: 10.21203/rs.3.rs-2229816/v1]
37. Raphadu TT, Staden MV, Dibakwane WM, Monyeki KD. A non-invasive investigation into the prevalence of higher than normal blood pressure, hypertension and the association between blood pressure and body weight in male and female adolescents in the Polokwane Local Municipality, Limpopo-South Africa: a cross-sectional study. Children (Basel) 2020 Mar 04;7(3):18 [FREE Full text] [doi: 10.3390/children7030018] [Medline: 32143272]




XSL•FO
RenderX

## Abbreviations

**AOR:** adjusted odds ratio
**HbA1c:** hemoglobin A1c
**HDL-C:** high-density lipoprotein cholesterol
**LDL-C:** low-density lipoprotein cholesterol
**LR:** logistic regression
**NCD:** noncommunicable disease
**SBP:** systolic blood pressure
**WC:** waist circumference
**WHO:** World Health Organization

*Edited by M Balcarras; submitted 19.Sep.2025; peer-reviewed by AA Lopez-Gonzalez, N Milinkovic; comments to author 15.May.2026; revised version received 29.Jun.2026; accepted 02.Jul.2026; published 10.Sep.2026.*

*Please cite as:*
*Thaga LF, Mukoma GG, Ellapen TJ, Muluvhu TC*
*Comparative Analysis of Cardiometabolic Risk Profiles Among Hypertensive and Normotensive Adults: Cross-Sectional Study*
*Interact J Med Res 2026;15:e84466*
*URL: https://www.i-jmr.org/2026/1/e84466*
*doi:10.2196/84466*
*PMID:*

©Lebogang Faith Thaga, Gudani Goodman Mukoma, Terry Jeremy Ellapen, Takalani Clearance Muluvhu. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 10.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# The Lifetime Dance Exposure Questionnaire for Professional Training: Survey Development and Reliability Study

Aston K McCullough[1,2,3], MA, MS, PhD; Kai Yi (Kaye) Han[1], BS

[1]Laboratory for the Scientific Study of Dance, Institute for Cognitive and Brain Health, Northeastern University, 360 Huntington Ave, 129 ISEC, Boston, MA, United States

[2]Department of Physical Therapy, Movement and Rehabilitation Sciences, Northeastern University, Boston, MA, United States

[3]Department of Music, Northeastern University, Boston, MA, United States

**Corresponding Author:**
Aston K McCullough, MA, MS, PhD
Laboratory for the Scientific Study of Dance, Institute for Cognitive and Brain Health, Northeastern University, 360 Huntington Ave, 129 ISEC, Boston, MA, United States

## Abstract

**Background:** Dance is a complex mode of physical activity (PA) behavior and an art form, and one's participation in dance may occur across discrete contexts throughout the lifespan. To further advance the study of lifetime participation in dance behavior, reliable questionnaires are needed.

**Objective:** This study aimed to develop and test the reliability of a self-report, online questionnaire for surveilling lifetime participation in professionally led dance classes among adults.

**Methods:** Community-dwelling adults (N=373) completed the Lifetime Dance Exposure Questionnaire for Professional Training (LDEQ-T) at baseline (T1), and 150 adults repeated the LDEQ-T after an 8-week delay (T2). Test-retest reliability for self-reported dance training frequency, duration, and PA intensity was analyzed for the LDEQ-T between T1 and T2. Reliability for the LDEQ-T item on total years of professional dance training was assessed using intraclass correlation coefficients (ICCs; 2-way, random effects); linearly weighted kappa ( $_w$) was used for ordinal variables on dance class duration, intensity, and frequency; an unweighted kappa ( ) statistic was used to test the reliability of 3 dance training exposure groups (no or low, homogenous, or variable lifetime exposure to professional dance training). The significance level was set a priori with $\alpha$=.05.

**Results:** Adults reported having no or low (n=92), homogenous (n=196), or variable (n=85) lifetime exposures to professional dance classes at T1. Among adults who self-reported homogenous exposures to dance training throughout the lifespan, years of dance training were found to be reliable across all age periods (ICC range 0.83-0.94); as were weeks/month, months/year, and hours/week of dance training ( $_w$ range 0.35-0.61); but not any measure of PA intensity. Among adults who self-reported variable exposures to dance training, years of dance training across all age periods (ICC=0.70-0.96); classes/week, weeks/month, months/year, and hours/week ( $_w$ range 0.32-0.56); and some estimates of PA intensity ( $_w$ range 0.30-0.41) were found to be sufficiently reliable.

**Conclusions:** The LDEQ-T appears to be a reliable instrument for quantifying lifetime participation in professional dance training among adults with self-reported low, homogenous, or variable exposures to professional dance classes throughout the lifespan. Among adults, researchers may surveil lifetime exposure to dance training in early childhood, childhood, adolescence, young adulthood, and middle or older adulthood using the LDEQ-T.

***(Interact J Med Res 2026;15:e75632)*** doi:10.2196/75632

**KEYWORDS**

survey; test-retest reliability; physical activity; self-report; adult

## Introduction

Dancing is positively associated with physiological, psychological, cognitive, and functional health outcomes across the lifespan [1-6]. Participation in dance refers to any of a wide range of unique or codified physical activity (PA) behaviors (eg, walking, running, sitting, standing, and jumping) performed with artistic or aesthetic intent [7]. It is widely known that PA dose—the product of the frequency, duration, and intensity of an exposure to a given mode of PA—is associated with health outcomes [8]. To that end, engaging in dance behavior, when compared to specific non-dance PA behaviors, has been reported to confer especial health benefits such as lower risk for cardiovascular disease mortality [9] and improve flexibility and balance [10]. However, additional research is needed to further extend available measures that are specifically designed to



XSL•FO
RenderX

surveil lifetime participation in dance behavior from a lifespan development perspective.

Some existing long-term PA recall surveys on lifetime PA participation include items on dance and provide general estimates of the total number of years spent engaged in dance [11,12], and a subset of these instruments provides self-reported estimates on the PA intensity of dance exposures throughout the lifespan [11]. Each of these instruments can be used to obtain global estimates of the total number of years spent participating in dance across solo, professional, and social contexts. However, studies show that the PA intensity of dance may depend, in part, on the conditions under which dancing is performed or upon participant-level factors. For example, a study on the PA intensity of modern dance across class, rehearsal, and performance contexts reported dance intensities that ranged from 2.9 metabolic equivalents (METs) to 7.1 METs depending upon the context in which the dance behavior was performed [13] Another study conducted in young to older adults reported the mean intensity of solo, free-form dancing with and without music ranged from 5.6 METs to 7.5 METs and that PA intensity during dance was inversely associated with age and BMI [14]. Given that the PA intensity of dance may vary when performed across different contexts, in addition to its inverse relationship with age, the need for self-report instruments that are expressly designed for quantifying context-specific exposures to dance across the lifespan is clear. Relatedly, studies show that lifetime participation in professionally led dance classes may be interrupted for variable periods of time across the lifespan due to injury [15], for example, or due to other barriers such as insufficient resources available to pay for dance classes [16]. Moreover, some research also indicates that early developmental exposures to professionally led dance classes may be negatively associated with psychological health outcomes during adulthood in some individuals [17]. Together, these studies point toward the need for context-specific measures of lifetime dance participation that allow researchers to specifically account for exposures to professionally led dance classes, while also accounting for potential variability in professional dance training exposures across the lifespan due to factors such as barriers to access.

Though some available measures of lifetime participation in PA provide the means for estimating years of dance participation generally (ie, undifferentiated exposures to dance across solo, professional, and social contexts), studies that aim to specifically quantify exposures to professionally led dance classes at different periods across the lifespan require the development of additional measures. Dance behavior, as a term, is inclusive of dance across any context (eg, solo dancing, dancing in rehearsal and performance, or dancing socially). Professionally led dance classes refer specifically to exposures to dance that are facilitated by a professional. With specific data on exposures to dance in professionally led class environments across the lifespan, researchers can better understand how the context in which dance behavior occurs relates to markers of health. Further, more granular data on professionally led dance classes with regard to chronological age and the timing of dance exposures across the lifespan may extend future analyses on the benefits of dance for health from a developmental perspective [18]. The design of a flexible instrument for quantifying exposures to professionally led dance classes across developmental periods would thus facilitate future research on dance and health from a lifespan development perspective and augment measures of lifetime participation in PA. Therefore, the purpose of this study was to develop and determine the reliability of a new PA questionnaire, the Lifetime Dance Exposure Questionnaire for Professional Training (LDEQ-T), specifically designed to surveil exposures to professionally led dance classes throughout the lifespan among adults.

## Methods

### Participants

Adults aged 18 years and older were recruited to participate in this research study. Prospective participants were recruited via flyers, social media (ie, Facebook [Meta] and Twitter), email, and word of mouth. Flyers were posted in the Pioneer Valley of Massachusetts and via social media. Within the Pioneer Valley, flyers were posted in dance studios, senior centers, at local community events, and within local businesses that were not directly associated with dance (eg, coffee shops or martial arts training venues). The link to the questionnaire was included in emails that the lab sent as part of a quarterly newsletter. Research team members shared information about the study informally throughout their personal networks and in social settings (ie, word of mouth).

### Procedures

Participants were asked to complete a brief sociodemographic questionnaire online via REDCap (Research Electronic Data Capture; Vanderbilt University) [19]. Participants were then asked to complete the novel LDEQ-T. After 8 weeks, participants received an email invitation via REDCap to complete the LDEQ-T for a second time. Participants received reminder emails to complete the LDEQ-T for 6 weeks or until they completed the questionnaire. Prior studies that have evaluated the reliability of self-report questionnaires for measuring lifetime exposures to PA have delayed readministering the instrument for intervals of 1 - 8 weeks [12,20,21], with some studies extending the delay period for up to 8 months [11].

### Sample Size Considerations

In considering the final study sample size, it was determined that at least thirty adults [22] should complete the LDEQ-T at the second time point.

### Materials

#### *Sociodemographic Questionnaire*

Self-reported age in years and biological sex were collected using a brief online questionnaire.

#### *LDEQ-T*

##### Development

The LDEQ-T is an original questionnaire designed by an expert (AKM) in dance, dance education, and movement science to



XSL•FO
RenderX

specifically quantify self-reported participation in professionally led dance classes throughout the lifespan.

### Conceptual Framework

As stated previously, PA dose is defined as the product of the intensity, duration, and frequency of a given PA exposure [8]. It follows that the LDEQ-T includes items on the PA intensity, duration, and frequency of professionally led dance classes taken throughout life.

### Item Generation and Draft Questionnaire

The questionnaire assesses total years of dance training, the duration, frequency, and intensity of dance exposures, and it categorizes respondents into 1 of 3 dance exposure groups based upon total dance training volume and regularity. The original questionnaire had a total of 24 items; however, the number of items differs for respondents based upon branching logic [23]. Respondents categorized as having low exposure complete a single-item questionnaire; respondents categorized as having homogenous exposure complete a 13-item questionnaire; respondents categorized as having variable exposure complete a 24-item questionnaire. Each exposure group is described further below.

#### *Exposure Groups*

Among those who report that they have completed ≥40 professionally led dance classes, the LDEQ-T accounts for potential variability in dance participation throughout the lifespan by asking respondents to identify whether there were any major discontinuities during their participation in professionally led dance classes throughout life. Individuals who self-report that they have completed <40 professionally led dance classes conclude the questionnaire once they have completed the first questionnaire item.

#### *Low Exposure*

A single item on the questionnaire (ie, "Have you participated in 40 or more professionally led dance classes in your life? For reference, participating in 40 dance classes is approximately equal to taking a dance class once a week, every week, for about 10 months.") was used to identify individuals with low exposure to professional dance training throughout life. The response format is a radio button [23] with options of "YES" and "NO" only. Respondents who self-report that they have completed <40 professionally led dance classes throughout life are categorized as individuals with low exposure to professionally led dance classes. The <40 class cut point was selected as a meaningful threshold that would represent having completed <1 year of dance training (ie, less than one class per week in a given year, assuming that annual curricula in a dance studio for children and adolescents may typically extend across a training period of ~10 months).

#### *Homogenous Exposure*

A single item on the questionnaire (ie, "Thinking only of the years you participated in dance classes on a regular basis, describe the usual amount of time that you spent taking dance classes year-to-year. Taking dance classes on a regular basis means participating in dance classes for 90 minutes or more per week, each week. While taking dance classes on a regular basis, the amount of time I spent taking dance classes year-to-year was.") was used to identify individuals whose exposure to professional dance training throughout life was consistent within and across dance training periods. The response format was a radio button with options of "about the SAME each year" and "much MORE on some years and much LESS on other years" only. Respondents who self-report engaging in dance classes consistently throughout life are categorized as individuals with homogenous exposure to professionally led dance classes.

#### *Variable Exposure*

Individuals who self-report discontinuities in dance class engagement throughout life (ie, individuals who reported taking dance classes "much MORE on some years and much LESS on other years") are categorized as individuals with variable exposures to professionally led dance classes.

### Total Years

The LDEQ-T asks respondents to identify each year in which they participated in dance classes that were led by a teacher from early childhood (age <2 years) to older adulthood (age >85 years). Using checkboxes [23], respondents are asked to select the checkbox next to each year that they engaged in professionally led dance classes. Individuals in the "Homogenous Exposure" group complete the total years item one time; those in the "Variable Exposure" group complete the total years item twice—once for the years in which they engaged in "MORE" dance classes than other years and once for the years in which they engaged "LESS" in taking dance classes. Questions on total years of participation were based upon the Lifetime Physical Activity Questionnaire [11], which asks respondents to select the year they began and stopped participating in each given mode of PA.

### Duration, Frequency, and Intensity

Finally, the questionnaire asks respondents to self-report the duration and PA intensity of the professionally led classes they attended. The format for these questions used drop-down menus and absolute open-ended quantifiers [23]. Individuals in the "Homogenous Exposure" group complete each item on the duration, intensity, and frequency one time; those in the "Variable Exposure" group complete each item on duration, intensity, and frequency twice—once for the years in which they engaged in "MORE" dance classes than other years and once for the years in which they engaged "LESS" in taking dance classes. Questions on duration, PA intensity, and frequency were based upon the Modifiable Activity Questionnaire and Godin-Shephard Leisure-Time PA questionnaire [24-26]. Questions on the duration and frequency of taking professionally led dance classes were fashioned after the Modifiable Activity Questionnaire, which asks respondents to report the number of times per month and the average number of minutes spent each time one engages in a given mode of PA. Questions on the perceived PA intensity of professionally led dance classes were fashioned after the Godin-Shephard Leisure-Time PA questionnaire, which asks respondents to report the total number of 15-minute bouts per week that they engaged in exercise at self-reported strenuous, moderate, and mild or light levels of effort. Examples of PA modes provided in the LDEQ-T for each level of effort were kept largely



XSL•FO
RenderX

consistent with those included in the original Godin-Shepard Leisure-Time PA questionnaire, with the inclusion of dance-related behaviors. LDEQ-T items on duration, intensity, and frequency appear below:

- Duration: hours per week spent taking dance classes (drop-down menu).
- Frequency: class sessions per week (absolute open-ended quantifier; min: 1; max: 40), weeks per month (drop-down menu), and months per year (drop-down menu).
- Intensity: (1) Minutes per class spent dancing at a strenuous effort level (absolute open-ended quantifier; min: 1; max: 240); (2) Minutes per class spent dancing at a moderate effort level (absolute open-ended quantifier; min: 1; max: 240); (3) Minutes per class spent dancing at a mild or light effort level (absolute open-ended quantifier; min: 1; max: 240); (4) Times per week spent dancing in class for about 15 minutes at a strenuous effort level (absolute open-ended quantifier; min: 1; max: 160); (5) Times per week spent dancing in class for about 15 minutes at a moderate effort level (absolute open-ended quantifier; min: 1; max: 160); and (6) Times per week spent dancing in class for about 15 minutes at a mild or light effort level (absolute open-ended quantifier; min: 1; max: 160).

### Statistical Analyses

Data were analyzed in MATLAB R2024a (The MathWorks, Inc). Descriptive statistics are reported as mean (SD), median (IQR), and frequencies (percentages). LDEQ-T test-retest reliability was evaluated using data collected at time point 1 (week 0) and time point 2 (week 8+). To do this, the item-level reliability of the LDEQ-T was evaluated across time points using intraclass correlation coefficients (ICCs) and Cohen kappa coefficients. ICCs (2-way, random effects) were used to evaluate whether LDEQ-T–derived estimates were sufficiently reliable (ICC≥0.70) for total years of dance training and years of training by age group (ie, early childhood, childhood, adolescence, young adulthood, and middle-age to older adulthood). Categories of low, moderate, and high exposure were respectively developed for the self-reported frequency (classes/wk, wk/mo, and mo/y) of taking professionally led dance classes and the duration (h/wk) of classes using tertiles. Similarly, tertiles were used to bin self-reported minutes spent engaged in each PA intensity while in class into categories of low, moderate, and high exposure. Finally, tertiles were used to bin the self-reported number of 15-minute bouts spent dancing in each PA intensity into categories of low, moderate, and high exposure. Reliability for each respective categorical measure was then determined using linearly weighted kappa coefficients ( $_w$) to account for the ordinal scale of the data, and unweighted kappa coefficients ( ) were reported otherwise [27,28].

To produce a final version of the LDEQ-T, any item that was found to have insufficient reliability was removed from the instrument following the item-level reliability analyses. The final version of the LDEQ-T can be scored simply by calculating the sum of the total years spent taking dance classes over any age period, and by applying the tertile-based cut points to any items on class duration, frequency, and intensity to categorize respondents into low, moderate, and high exposure groups for each respective item. The questionnaire may also be scored with the LDEQ-T companion app [29], which was developed by the researcher (AKM) using MATLAB and is available to be downloaded. The significance level in all models was set a priori with $\alpha$=.05.

### Ethical Considerations

The study was approved by the Institutional Review Board at the University of Massachusetts Amherst (IRB #2495), and all participants provided written informed consent. Data were collected between February 2022 and August 2023, data were deidentified, and the study was conducted entirely online.

## Results

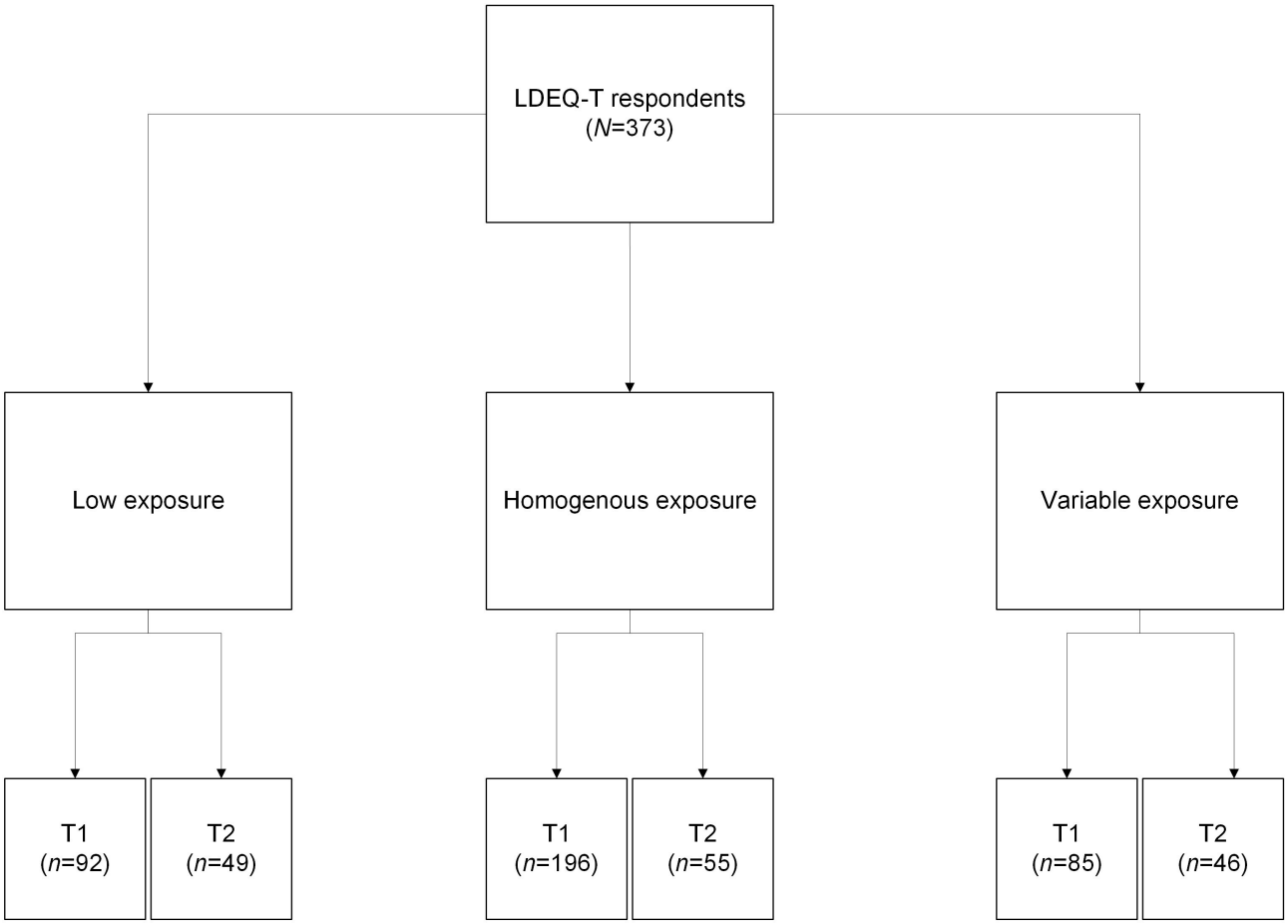
### Overview

A total of 373 adults (Figure 1) completed the LDEQ-T at timepoint one (T1). Of the overall sample, 365 adults reported their age (range 18 - 84 y) at T1; 81.9% (n=299) were young adults (aged 18 - 39 years; median 27, IQR 23-30) and 18.1% (n=66) were middle age to older adults (aged 40 - 84 y; median 49, IQR 43-57). A total of 150 adults repeated the questionnaire at time point 2 (T2). LDEQ-T test-retest reliability between T1 and T2 was sufficient for the self-reported lifetime exposure categories of low exposure, homogenous exposure, and variable exposure [ (SE)=0.53 (0.06), 95% CI 0.42-0.64; *P*<.001]. Figure 2A shows self-reported dance training across the lifespan, by biological sex, using LDEQ-T responses collected at T1. The range of self-reported years of participation in professionally led dance classes ranged from 0 to 68 total years at T1.



XSL•FO
RenderX

**Figure 1.** Responses to the Lifetime Dance Exposure Questionnaire for Professional Training note: Figure 1 shows the total number of Lifetime Dance Exposure Questionnaire for Professional Training responses collected at baseline (timepoint 1; T1) and after an 8+ week delay (timepoint 2; T2) among adults who self-reported having low exposure to professionally led dance classes throughout life, in addition to those who reported homogenous or variable exposures to dance training year-to-year. LDEQ-T: Lifetime Dance Exposure Questionnaire for Professional Training.






XSL•FO
RenderX

**Figure 2.** Participation in professional dance classes across the lifespan by sex (A) and by dance exposure pattern (B) among adults aged 18 to 84 years as reported using the Lifetime Dance Exposure Questionnaire (LDEQ-T). Note: Figure 2A shows the relative percentage of adults who self-reported participating in professionally led dance classes across the lifespan for female and male LDEQ-T respondents. Figure 2B shows the relative percentage of adults who self-reported participating in professionally led dance classes across the lifespan by exposure pattern (ie, homogenous or variable exposures to dance).



XSL•FO
RenderX

### Low Exposure to Dance Training

A total of 92 adults (mean age 32, SD 12.8 y; biological sex: female 52.2%, 48/92, male 47.8%, 44/92) self-reported engaging in <40 professionally led dance classes throughout life, and of these, a total of 49 adults completed the LDEQ-T at T2. The percent agreement between T1 and T2 for those who self-reported low exposure to dance training was 73.5%—among adults who reported low exposure to dance training at T1, the range of self-reported total years of dance training at T2 was 0 to 16 years.

### Homogenous Exposure to Dance Training

A total of 196 adults (mean age 30, SD 10.7 y; biological sex: female 60.7%, 119/196, male 38.8%, 76/196, not reported 0.5%, 1/196) indicated that they experienced homogenous exposures to professionally led dance classes throughout life when completing the LDEQ-T at T1, and of these n=55 adults completed the LDEQ-T at T2. Figure 2B shows self-reported exposures to professionally led dance classes across the lifespan for adults who reported homogenous exposures to dance training at T1.

Self-reported years of dance training were found to be sufficiently reliable (ICC range 0.83-0.94) for those who self-reported experiencing homogenous exposures to professionally led dance classes across the lifespan (Table 1). As Table 2 shows, the frequency and duration of self-reported professional dance classes were also found to be reliable when assessed in terms of weeks/month, months/year, and hours/week ( $_w$ range 0.35-0.61). Remaining items on frequency, duration, and intensity had low kappa coefficients and were excluded from the final version of the LDEQ-T due to insufficient reliability ($P$>.05). These items included average class sessions per week ( $_w$=0.23); average minutes of dance at strenuous ( $_w$=0.18), moderate ( $_w$=0.19), and mild or light levels of perceived exertion ( $_w$=0.17) per class; and average times per week spent dancing for about 15 minutes in class at strenuous ( $_w$=0.21), moderate ( $_w$=0.10), and mild or light ( $_w$=0.05) levels of perceived exertion.

**Table .** Lifetime Dance Exposure Questionnaire for Professional Training reliability (homogenous exposure): years of training by age group. Note: Table 1 shows the mean (SD), and intraclass correlation coefficients for years of dance training for five age groups across timepoint 1 (T1) and timepoint 2 (T2) among adults who reported having homogenous exposures to professionally led dance classes throughout life.

| Years of training | T1, mean (SD) | T2, mean (SD) | ICC[a] (95% CI) |
|---|---|---|---|
| Total lifetime training (n=55) | 7.3 (8.9) | 6.7 (8.5) | 0.89 (0.85-0.93) |
| Early childhood training (age >2 to 5 years) | 0.5 (1) | 0.5 (1.1) | 0.81 (0.75-0.86) |
| Childhood training (age 6 to 11 years) | 1.6 (2.1) | 1.7 (2.2) | 0.91 (0.89-0.93) |
| Training during adolescence (age 12 to 17 years) | 1.6 (2.1) | 1.7 (2.3) | 0.89 (0.85-0.93) |
| Training during young adulthood (age 18 to 39 years) | 2.5 (4.8) | 2.4 (4.6) | 0.94 (0.92-0.95) |
| Training during middle and older adulthood (age >40 years; n=28) | 0.8 (2.5) | 0.5 (1.8) | 0.83 (0.73-0.93) |

[a]ICC: intraclass correlation coefficients.



XSL•FO
RenderX

**Table .** Lifetime Dance Exposure Questionnaire for Professional Training reliability (homogenous dance exposures): Dance class frequency and duration tertiles. Note: Table 2 summarizes the percentage of participants within low, moderate, and high exposure tertiles for LDEQ-T items on dance exposure frequency (weeks/months; months/year) and duration (hours/week) at baseline (T1) and after an 8+ week delay (T2). Italicized numbers show corresponding tertiles for each respective item on frequency and duration.

| Measures | Low, n (%) | Moderate, n (%) | High, n (%) | $_w$ | SE (95% CI) | *P* value |
|---|---|---|---|---|---|---|
| Weeks per month (n=55) | *1-2* | *3* | *4* | 0.61 | 0.11 (0.40-0.82) | <.001 |
| T1 | 18 (32.7) | 5 (9.1) | 32 (58.2) | | | |
| T2 | 24 (43.6) | 4 (7.3) | 27 (49.1) | | | |
| Months per year (n=55) | *1-5* | *6-11* | *12* | 0.35 | 0.14 (0.09-0.62) | .009 |
| T1 | 14 (25.4) | 21 (38.2) | 20 (36.4) | | | |
| T2 | 26 (47.3) | 19 (34.5) | 10 (18.2) | | | |
| Hours per week (n=55) | *1-3* | *4-10* | *>10* | 0.38 | 0.15 (0.09-0.67) | .01 |
| T1 | 22 (40) | 22 (40) | 11 (20) | | | |
| T2 | 35 (64.8) | 13 (24.1) | 6 (11.1) | | | |

## Variable Exposure to Dance Training

A total of 85 participants completed the LDEQ-T at T1 and indicated having variable exposure to dance training throughout life (mean age 30, SD 13.4 y; biological sex: female 68.2%, 58/85, male 30.6%, 26/85, not reported 1.2%, 1/85), and of these, a total of 46 adults completed the LDEQ-T at T2. Figure 2B shows self-reported dance training across the lifespan for adults who reported variable exposure to dance training at T1. Self-reported years of dance training were found to be sufficiently reliable (ICC range 0.70-0.96) for those who reported experiencing variable exposures to dance training across the lifespan (Table 3).

**Table .** Lifetime Dance Exposure Questionnaire for Professional Training reliability (variable dance exposure): Years of training by age group. Note: Table 3 shows the mean (SD), and intraclass correlation coefficients (ICC) for years of dance training for five age groups across timepoint 1 (T1) and timepoint 2 (T2) among adults who reported having variable exposure to professionally led dance classes throughout life.

| Years of training | T1, mean (SD) | T2, mean (SD) | ICC[a] (95% CI) |
|---|---|---|---|
| Total lifetime training (n=46) | 13.6 (13.1) | 11.6 (10.4) | 0.93 (0.89-0.96) |
| Early childhood training (age >2 to 5 years) | 0.8 (1.2) | 0.7 (1.2) | 0.80 (0.76-0.84) |
| Childhood training (age 6 to 11 years) | 2.8 (2.6) | 2.6 (2.6) | 0.90 (0.88-0.92) |
| Training during adolescence (age 12 to 17 years) | 3 (2.6) | 2.8 (2.6) | 0.93 (0.90-0.95) |
| Training during young adulthood (age 18 to 39 years) | 5.6 (6.2) | 5 (5.7) | 0.96 (0.94-0.97) |
| Training during middle and older adulthood (age >40 years; n=18) | 1.2 (4.4) | 0.5 (1.8) | 0.70 (0.56-0.84) |

[a]ICC: intraclass correlation coefficients.

### *High Exposure Years*

As shown in Table 4, the frequency and duration of self-reported professional dance classes were also found to be reliable when assessed in terms of classes/week, weeks/month, months/year, and hours/week ( $_w$ range 0.38-0.56).



XSL•FO
RenderX

**Table .** Lifetime Dance Exposure Questionnaire for Professional Training reliability (variable dance exposure): Dance class frequency and duration tertiles during high exposure years. Note: Table 4 summarizes the percentage of participants within low, moderate, and high exposure tertiles for items on dance exposure frequency (weeks/months; months/year) and duration (hours/week) during their high dance exposure years at baseline (T1) and after an 8+ week delay (T2). Italic values represent corresponding tertiles for each respective item on frequency and duration.

| Measures | Low | Moderate | High | $_w$ | SE (95% CI) | *P* value |
|---|---|---|---|---|---|---|
| Classes per week (N=46) | *1-2* | *3-5* | *>5* | 0.56 | 0.13 (0.31-0.81) | <.001 |
| T1 | 7 (15.2) | 15 (32.6) | 24 (52.2) | | | |
| T2 | 19 (41.3) | 8 (17.4) | 19 (41.3) | | | |
| Weeks per month (N=46) | *1-2* | *3* | *4* | 0.38 | 0.16 (0.07-0.69) | .016 |
| T1 | 5 (10.9) | 6 (13) | 35 (76.1) | | | |
| T2 | 15 (32.6) | 3 (6.5) | 28 (60.9) | | | |
| Months per year (N=46) | *1-8* | *9-11* | *12* | 0.47 | 0.14 (0.20-0.74) | <.001 |
| T1 | 8 (17.4) | 20 (43.5) | 18 (39.1) | | | |
| T2 | 20 (43.5) | 14 (30.4) | 12 (26.1) | | | |
| Hours per week (N=46) | *1-5* | *6-13* | *>13* | 0.55 | 0.13 (0.30-0.81) | <.001 |
| T1 | 16 (34.8) | 15 (32.6) | 15 (32.6) | | | |
| T2 | 22 (47.8) | 10 (21.8) | 14 (30.4) | | | |

All but 2 items pertaining to dance intensity were reliable, summarized in Table 5 ( $_w$: 0.33-0.42). Items with low kappa coefficients and insufficient reliability (*P*>.05) were average class minutes spent dancing at mild or light levels of perceived exertion ( $_w$=0.17) and average times per week spent dancing for about 15 minutes in class at strenuous levels of perceived exertion ( $_w$=0.23), which were excluded from the finalized LDEQ-T.

**Table .** Lifetime Dance Exposure Questionnaire for Professional Training reliability (variable dance exposure): Time spent engaged in dance class at mild to strenuous physical activity intensities—tertiles during high exposure years. Note: Table 5 summarizes the percentage of participants within low, moderate, and high tertiles for items on dance intensity (minutes and number of 15-minute bouts of dancing per week at different levels of perceived exertion) during their high dance exposure years at baseline (T1) and after an 8+ week delay (T2). Italic values represent corresponding tertiles for each respective item on intensity.

| Measures | Low | Moderate | High | $\kappa_w$ | SE (95% CI) | *P* value |
|---|---|---|---|---|---|---|
| Strenuous minutes (n=46) | *1 - 19* | *20 - 33* | *>33* | 0.33 | 0.15 (0.04-0.62) | .027 |
| T1 | 11 (23.9) | 16 (34.8) | 19 (41.3) | | | |
| T2 | 21 (46.7) | 14 (31.1) | 10 (22.2) | | | |
| Moderate minutes (n=46) | *1 - 19* | *20 - 33* | *>33* | 0.41 | 0.14 (0.13-0.68) | .004 |
| T1 | 12 (26.1) | 14 (30.4) | 20 (43.5) | | | |
| T2 | 21 (45.7) | 12 (26.1) | 13 (28.2) | | | |
| Moderate bouts (n=46) | *1 - 2* | *3 - 6* | *>6* | 0.36 | 0.15 (0.06-0.65) | .019 |
| T1 | 13 (28.3) | 17 (36.9) | 16 (34.8) | | | |
| T2 | 19 (41.3) | 17 (37) | 10 (21.7) | | | |
| Mild or light bouts (n=46) | *1 - 2* | *3 - 6* | *>6* | 0.42 | 0.15 (0.13-0.71) | .005 |
| T1 | 14 (30.43) | 18 (39.13) | 14 (30.44) | | | |
| T2 | 19 (41.3) | 17 (37) | 10 (21.7) | | | |



XSL•FO
RenderX

### *Low Exposure Years*

As shown in Table 6, the frequency and duration of self-reported professional dance classes were also found to be reliable when assessed in terms of classes/week, weeks/month, months/year, and hours/week ( $_w$ range 0.32-0.43) during self-reported less active dance training years.

**Table .** Lifetime Dance Exposure Questionnaire for Professional Training reliability (variable dance exposure): Dance class frequency and duration tertiles during low exposure years. Note: Table 6 summarizes the percentage of participants within low, moderate, and high exposure tertiles for items on dance exposure frequency (weeks/months; months/year) and duration (hours/week) during their low dance exposure years at baseline (T1) and after an 8+ week delay (T2). Italic values represent corresponding tertiles for each respective item on frequency and duration.

| Measures | Low | Moderate | High | $_w$ | SE (95% CI) | *P* value |
|---|---|---|---|---|---|---|
| Classes per week (N=46) | *1* | *2* | *3* | 0.33 | 0.15 (0.06-0.62) | .021 |
| T1 | 15 (32.6) | 13 (28.3) | 18 (31.1) | | | |
| T2 | 19 (45.7) | 10 (21.7) | 15 (32.6) | | | |
| Weeks per month (N=46) | *1* | *2-3* | *>3* | 0.33 | 0.15 (0.03-0.63) | .033 |
| T1 | 5 (10.8) | 13 (28.3) | 28 (60.9) | | | |
| T2 | 15 (32.6) | 10 (21.7) | 21 (45.7) | | | |
| Months per year (N=46) | *1-5* | *6-10* | *>10* | 0.43 | 0.14 (0.16-0.70) | .002 |
| T1 | 9 (19.6) | 7 (15.2) | 30 (65.2) | | | |
| T2 | 19 (41.3) | 4 (8.7) | 23 (50) | | | |
| Hours per week (N=46) | *1-2* | *3-6* | *>6* | 0.32 | 0.16 (0.01-0.62) | .043 |
| T1 | 17 (37) | 15 (32.6) | 14 (30.4) | | | |
| T2 | 25 (54.3) | 16 (34.8) | 5 (10.9) | | | |

Two items on perceived dance intensity were sufficiently reliable for less active dance training years, as summarized in Table 7 ( $_w$ range 0.30-0.31). Remaining items with low kappa coefficients and insufficient reliability (*P*>.05) were excluded from the final LDEQ-T, which were average class minutes spent dancing at strenuous levels of perceived exertion ( $_w$=0.30) and average times per week spent dancing for about 15 minutes in class at strenuous ( $_w$=0.23), moderate ( $_w$=0.24), and mild or light ( $_w$=0.21) levels of perceived exertion.

**Table .** Lifetime Dance Exposure Questionnaire for Professional Training reliability (variable dance exposure): Minutes spent engaged in dance class at mild or light and moderate physical activity intensities—tertiles during low exposure years. Note: Table 7 summarizes the percentage of participants within low, moderate, and high tertiles for items on dance intensity (minutes of dancing per week at different levels of perceived exertion) during their low dance exposure years at baseline (T1) and after an 8+ week delay (T2). Italic values represent corresponding tertiles for each respective item on intensity.

| Measures | Low | Moderate | High | $_w$ | SE (95% CI) | *P* value |
|---|---|---|---|---|---|---|
| Moderate minutes (N=46) | *1 - 14* | *15 - 33* | *>33* | 0.30 | 0.15 (0.001-0.598) | .049 |
| T1 | 9 (19.6) | 20 (43.5) | 17 (36.9) | | | |
| T2 | 21 (45.7) | 15 (32.6) | 10 (21.7) | | | |
| Mild or light minutes (N=46) | *1 - 14* | *15 - 23* | *>23* | 0.31 | 0.14 (0.03-0.59) | .031 |
| T1 | 12 (26.1) | 15 (32.6) | 19 (41.3) | | | |
| T2 | 25 (54.4) | 10 (21.7) | 11 (23.9) | | | |

## LDEQ-T Refinement and Companion App

### *Questionnaire Refinement*

Items with insufficient reliability were removed from the original 24-item version of LDEQ-T. The final LDEQ-T (18 items) can be downloaded as a *.pdf and as a REDCap "Data Dictionary" [29]. The total number of items for each exposure group in the final version of the questionnaire was 1, 6, and 18 for those in the low, homogenous, and variable exposure groups, respectively. Field notes show that the questionnaire may be completed in 1 - 15 minutes.



XSL•FO
RenderX

### *LDEQ-T Companion App*

The LDEQ-T app (KinetiWave) allows users to select a spreadsheet with LDEQ-T responses, as exported from REDCap, and it returns a summary of the total number of years spent engaged in professional dance training, as well as a summary of the years spent engaged in dance classes across each age group from early childhood to middle and older adulthood. In addition, the LDEQ-T companion app applies the empirically derived, tertile-based cutoff points to all LDEQ-T items on the self-reported duration, frequency, and perceived intensity of professionally led dance classes. An example summary report can be downloaded with the LDEQ-T app [29].

## Discussion

### Overview

This study aimed to determine the reliability of a new questionnaire, the LDEQ-T, which was specifically designed to surveil lifetime exposures to professionally led dance classes among adults. Analyses of the LDEQ-T revealed that it is a reliable questionnaire for assessing lifetime participation in professionally led dance classes among adults aged 18 to 84 years. When the LDEQ-T was used to assess dance training among adults who self-reported experiencing homogenous exposures to professionally led dance classes throughout life, self-reports on the frequency and duration of dance classes, but not the PA intensity of classes, were found to be reliable. Among adults who self-reported experiencing variable exposure to professionally led dance classes across life, self-reports on the frequency, duration, and the PA intensity of dance classes were respectively found to be reliable for LDEQ-T items depending upon the overall level of self-reported dance training activity each year. Overall, results show the LDEQ-T is a reliable instrument for assessing lifetime participation in professionally led dance classes among adults with a range of prior exposures to dance training.

### Principal Results

In a sample of adults aged 18 to 84 years, the LDEQ-T provided reliable estimates of the total number of years in which respondents engaged in professionally led dance classes across the lifespan. Among a cohort of adults with prior exposures to dance training, LDEQ-T reliability estimates were observed to be good to very good (ICC>0.89) for self-reported total years of experience in professionally led dance classes. Reliability estimates for the LDEQ-T were also observed to be sufficient when years of dance training were analyzed by dance exposure by age groups, early childhood, adolescence, young adulthood, middle or older adulthood (ICC>0.70). Study results showed that the LDEQ-T can be used as a reliable measure of the total number of years in which adults have engaged in professionally led dance classes, in addition to providing a measure of age-specific exposures to dance training across the lifespan. Prior studies that have evaluated the reliability of lifetime PA questionnaires report ICCs in the range 0.67 - 0.89 [11,12]. Thus, the reliability of the LDEQ-T is comparatively strong, and future studies may use the LDEQ-T to obtain self-reports of the total number of years one has engaged in professionally led dance classes across the lifespan, as well as assess exposure to professionally led dance classes by developmental age groups.

Reliability for LDEQ-T items about the PA intensity of dance classes was found to be sufficient among adults who self-reported experiencing variable exposure to dance training throughout life. Specifically, during years that adults reported as being more active dance training years, reliability was found to be sufficiently high for items on self-reported time spent at strenuous and moderate ( $_w$: 0.33-0.41) levels of perceived effort and on self-reported weekly bouts of activity performed at moderate and mild or light ( $_w$: 0.36-0.42) levels of perceived effort. Similarly, for the years adults reported they were less engaged in dance training, reliability was found to be sufficient [30] for items on self-reported minutes spent at moderate and mild or light levels of perceived effort ( $_w$: 0.30-0.31). For adults who reported experiencing homogenous exposures to dance training throughout life, no items on the perceived intensity of dance classes were found to be sufficiently reliable. Average years of dance training experience were qualitatively higher among adults with variable exposures to dance training (13.6 y) versus those with homogenous exposures to dance training (7.3 y). It is possible that those who participated in dance over a greater number of years were more consistently able to recall the perceived PA intensity of professionally led dance classes; however, additional research is needed to understand how lifetime PA exposure patterns may potentially affect long-term PA intensity recall.

The results of this study suggest that, among adults with variable exposures to dance training throughout life, the LDEQ-T can be used to recall and assess the self-reported perceived PA intensity of dance classes during both more active and less active dance training years. Among adults with homogenous exposures to dance training, results also show the LDEQ-T can be used to recall and assess the self-reported frequency and duration of dance classes taken throughout life. The LDEQ-T questionnaire can be obtained, along with a copy of the questionnaire items, response scale characteristics, and REDCap “Data Dictionary” as part of the LDEQ-T app download [29]. The LDEQ-T app includes a calculator for LDEQ-T responses; the app applies the tertile cut points, as presented in our study results, and returns summary statistics on professionally led dance class exposure for respondents. Work from our group showed that total years of exposure to professionally led dance classes, as measured by the LDEQ-T, were related to movement qualities associated with free-form dance in a study that used wearable devices to quantify motor behavior quality among adults aged 18 to 83 years with a range of 0 to 56 years of professional dance training [31].

### Limitations and Delimitations

A limitation of the study is that no concurrent measure of lifetime exposure to professionally led dance training was used to validate the LDEQ-T. This is because, to our knowledge, this is the first study to develop a method for surveilling exposure to professionally led dance classes throughout life. While the lack of a concurrent measure of PA may be a limitation that is typical of most studies that have developed long-term PA recall questionnaires [21,32], studies have also been previously



XSL•FO
RenderX

conducted that concurrently integrated the use of wearable devices in the development of long-term PA recall instruments [33]. Importantly, a major difference between this study and prior studies that have used wearable devices to validate long-term PA recall instruments is that the LDEQ-T accounts for dance training in childhood, while prior studies that have used wearable devices to validate their questionnaire were focused only on PA in adulthood. Future studies on dance that seek to further develop surveys for long-term PA recall should consider including wearable sensors as part of the questionnaire design process to further advance PA recall-based methods [32,33]. Given that no other measure exists for expressly assessing exposures to professionally led dance classes across the lifespan, all reports on years of dance training experience that were represented in this study sample were obtained from the LDEQ-T. A delimitation of the study is that children and adolescents were not included in the sample. The study sample included young to older adults, and the item-level reliability of the LDEQ-T questionnaire was found to be sufficient across age groups. Nevertheless, it is important to note that the study sample was comprised mostly of adults aged 18 to 39 years (~82%), and adults aged 40 to 84 years comprised the remaining ~18 % of the sample. Future studies are required to determine if the LDEQ-T can be used among non-adult respondents; additional research on the LDEQ-T among older adults, especially >75 years, may be needed. Finally, the LDEQ-T does not include any open-ended questions on self-reported dance style; therefore, researchers may need to additionally allow respondents to provide qualitative feedback in order to further document dance heritage or style when characterizing lifetime dance exposure.

### Strengths

A strength of the study is that a sample of young to older adults, aged 18 to 84 years, both with and without prior dance training, were recruited to complete the LDEQ-T. The LDEQ-T provides a novel measure of dance training that allows researchers to account for variable exposures to dance training throughout the lifespan. Furthermore, the LDEQ-T offers a context-specific measure of dance training that focuses on exposure to dance in professionally led classes. In view of the promising findings presented in this study, we recommend that researchers consider using the LDEQ-T, and its companion calculator app [29], to assess exposures to professionally led dance classes across the lifespan in future studies of dance and health among adults with or without prior dance training. Finally, we present the item-level reliability for the LDEQ-T questionnaire, which enables researchers to select and use any of the LDEQ-T items that are most relevant to the purposes of their research.

### Conclusions

The LDEQ-T appears to be a reliable self-report instrument for assessing the frequency and duration of lifetime participation in professionally led dance classes among adults aged 18 to 84 years, with a range of 0 to 68 years of dance training. Adults who have experienced variable exposure to professional dance training throughout life may additionally use the LDEQ-T to reliably report the perceived intensity of dance classes. Future studies on lifetime exposures to dance training among adults may use the LDEQ-T to surveil dance exposures from a lifespan development perspective.

## Acknowledgments

The authors wish to thank the study participants who volunteered their time, and the research team members from Laboratory for the Scientific Study of Dance (LAB:SYNC) for their diverse support. The authors also wish to thank the specialists in dance and physical activity who provided feedback on the initial draft of the Lifetime Dance Exposure Questionnaire for Professional Training. Disclaimer: The opinions expressed are those of the author and do not represent the views of the National Endowment for the Arts (NEA) Office of Research & Analysis or the National Endowment for the Arts. The NEA does not guarantee the accuracy or completeness of the information included in this material and is not responsible for any consequences of its use.

## Funding

The study was funded in part by an award from the National Endowment for the Arts(Award #: 1879058-38-C-21).

## Data Availability

Study data presented in this article may be obtained upon submitting a request to the corresponding author.

## Authors' Contributions

Conceptualization: AKM
Methodology: AKM
Software: AKM
Validation: AKM
Formal analysis: AKM
Investigation: AKM
Resources: AKM
Data curation: AKM
Writing – initial draft: AKM



XSL•FO
RenderX

Writing – Revision: AKM and KH
Visualization: AKM and KH
Supervision: AKM
Project administration: AKM
Funding acquisition: AKM

## Conflicts of Interest

None declared.

## References


1. Clifford AM, Shanahan J, McKee J, et al. The effect of dance on physical health and cognition in community dwelling older adults: a systematic review and meta-analysis. Arts Health 2023 Jun;15(2):200-228. [doi: 10.1080/17533015.2022.2093929] [Medline: 35969880]
2. Fong Yan A, Cobley S, Chan C, et al. The effectiveness of dance interventions on physical health outcomes compared to other forms of physical activity: a systematic review and meta-analysis. Sports Med 2018 Apr;48(4):933-951. [doi: 10.1007/s40279-017-0853-5] [Medline: 29270864]
3. Fong Yan A, Nicholson LL, Ward RE, et al. The effectiveness of dance interventions on psychological and cognitive health outcomes compared with other forms of physical activity: a systematic review with meta-analysis. Sports Med 2024 May;54(5):1179-1205. [doi: 10.1007/s40279-023-01990-2] [Medline: 38270792]
4. Aldemir GY, Ramazano lu N, Çamligüney AF, et al. The effects of dance education on motor performance of children. Educ Res Rev 2011;6:979-982 [FREE Full text] [doi: 10.5897/ERR11.179]
5. Wu VX, Chi Y, Lee JK, et al. The effect of dance interventions on cognition, neuroplasticity, physical function, depression, and quality of life for older adults with mild cognitive impairment: a systematic review and meta-analysis. Int J Nurs Stud 2021 Oct;122:104025. [doi: 10.1016/j.ijnurstu.2021.104025] [Medline: 34298320]
6. Mattle M, Chocano-Bedoya PO, Fischbacher M, et al. Association of dance-based mind-motor activities with falls and physical function among healthy older adults: a systematic review and meta-analysis. JAMA Netw Open 2020 Sep 1;3(9):e2017688. [doi: 10.1001/jamanetworkopen.2020.17688] [Medline: 32975570]
7. Hawkins AM. Creating Through Dance: Prentice-Hall; 1964. URL: https://archive.org/details/creatingthroughd0000hawk [accessed 2026-03-18]
8. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985;100(2):126-131. [Medline: 3920711]
9. Merom D, Ding D, Stamatakis E. Dancing participation and cardiovascular disease mortality: a pooled analysis of 11 population-based British cohorts. Am J Prev Med 2016 Jun;50(6):756-760. [doi: 10.1016/j.amepre.2016.01.004] [Medline: 26944521]
10. Lu J, Abd Rahman NA, Wyon M, Shaharudin S. The effects of dance interventions on physical function and quality of life among middle-aged and older adults: a systematic review. PLoS ONE 2024;19(4):e0301236. [doi: 10.1371/journal.pone.0301236] [Medline: 38640093]
11. De Vera MA, Ratzlaff C, Doerfling P, Kopec J. Reliability and validity of an internet-based questionnaire measuring lifetime physical activity. Am J Epidemiol 2010 Nov 15;172(10):1190-1198. [doi: 10.1093/aje/kwq273] [Medline: 20876666]
12. Chasan-Taber L, Erickson JB, McBride JW, Nasca PC, Chasan-Taber S, Freedson PS. Reproducibility of a self-administered lifetime physical activity questionnaire among female college alumnae. Am J Epidemiol 2002 Feb 1;155(3):282-289. [doi: 10.1093/aje/155.3.282] [Medline: 11821254]
13. Wyon MA, Abt G, Redding E, Head A, Sharp NCC. Oxygen uptake during modern dance class, rehearsal, and performance. J Strength Cond Res 2004 Aug;18(3):646-649. [doi: 10.1519/13082.1] [Medline: 15320648]
14. McCullough AK. Absolute and relative intensities of solo, free-form dancing in adults: a pilot study. PLoS ONE 2024;19(11):e0313144. [doi: 10.1371/journal.pone.0313144] [Medline: 39556599]
15. Harrison C, Ruddock-Hudson M. Perceptions of pain, injury, and transition-retirement: the experiences of professional dancers. J Dance Med Sci 2017 Jun 15;21(2):43-52. [doi: 10.12678/1089-313X.21.2.43] [Medline: 28535847]
16. Lawson HM. Why dance? The motivations of an unlikely group of dancers. MAiA 2009;1:3-15 [FREE Full text]
17. Ackard DM, Henderson JB, Wonderlich AL. The associations between childhood dance participation and adult disordered eating and related psychopathology. J Psychosom Res 2004 Nov;57(5):485-490. [doi: 10.1016/j.jpsychores.2004.03.004] [Medline: 15581653]
18. Geifman N, Cohen R, Rubin E. Redefining meaningful age groups in the context of disease. Age (Dordr) 2013 Dec;35(6):2357-2366. [doi: 10.1007/s11357-013-9510-6] [Medline: 23354682]
19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]




XSL•FO
RenderX

20. Kriska AM, Sandler RB, Cauley JA, LaPorte RE, Hom DL, Pambianco G. The assessment of historical physical activity and its relation to adult bone parameters. Am J Epidemiol 1988 May;127(5):1053-1063. [doi: 10.1093/oxfordjournals.aje.a114881] [Medline: 3358406]
21. Friedenreich CM, Courneya KS, Bryant HE. The lifetime total physical activity questionnaire: development and reliability. Med Sci Sports Exerc 1998 Feb;30(2):266-274. [doi: 10.1097/00005768-199802000-00015] [Medline: 9502356]
22. Bujang MA, Omar ED, Foo DHP, Hon YK. Sample size determination for conducting a pilot study to assess reliability of a questionnaire. Restor Dent Endod 2024 Feb;49(1):e3. [doi: 10.5395/rde.2024.49.e3] [Medline: 38449496]
23. DeCastellarnau A. A classification of response scale characteristics that affect data quality: a literature review. Qual Quant 2018;52(4):1523-1559. [doi: 10.1007/s11135-017-0533-4] [Medline: 29937582]
24. Amireault S, Godin G. The Godin-Shephard leisure-time physical activity questionnaire: validity evidence supporting its use for classifying healthy adults into active and insufficiently active categories. Percept Mot Skills 2015 Apr;120(2):604-622. [doi: 10.2466/03.27.PMS.120v19x7] [Medline: 25799030]
25. Modifiable activity questionnaire. Med Sci Sports Exerc 1997 Jun;29(Supplement):73-78. [doi: 10.1097/00005768-199706001-00015]
26. Pettee Gabriel K, McClain JJ, Schmid KK, Storti KL, Ainsworth BE. Reliability and convergent validity of the past-week modifiable activity questionnaire. Public Health Nutr 2011 Mar;14(3):435-442. [doi: 10.1017/S1368980010002612] [Medline: 20843404]
27. Vanbelle S, Albert A. A note on the linearly weighted kappa coefficient for ordinal scales. Stat Methodol 2009 Mar;6(2):157-163. [doi: 10.1016/j.stamet.2008.06.001]
28. Cardillo G. Cohen's kappa. GitHub. 2025. URL: https://www.mathworks.com/matlabcentral/fileexchange/15365-cohen-s-kappa [accessed 2024-12-09]
29. McCullough AK. Lifetime dance exposure questionnaire (LDEQ). : KinetiWave; 2025 Apr. [doi: 10.5281/zenodo.15171302]
30. Brown WJ, Trost SG, Bauman A, Mummery K, Owen N. Test-retest reliability of four physical activity measures used in population surveys. J Sci Med Sport 2004 Jun;7(2):205-215. [doi: 10.1016/s1440-2440(04)80010-0] [Medline: 15362316]
31. McCullough AK. Use of posterior probabilities from a long short-term memory network for characterizing dance behavior with multiple accelerometers. J Alzheimers Dis 2025 Jun;105(4):1069-1084. [doi: 10.1177/13872877251336482] [Medline: 40320775]
32. Smith AW, Cronin KA, Bowles H, et al. Reproducibility of physical activity recall over fifteen years: longitudinal evidence from the CARDIA study. BMC Public Health 2013 Feb 28;13:180. [doi: 10.1186/1471-2458-13-180] [Medline: 23448132]
33. Besson H, Harwood CA, Ekelund U, et al. Validation of the historical adulthood physical activity questionnaire (HAPAQ) against objective measurements of physical activity. Int J Behav Nutr Phys Act 2010 Jun 24;7:54. [doi: 10.1186/1479-5868-7-54] [Medline: 20576086]


## Abbreviations

**ICC:** intraclass correlation coefficients
**LDEQ-T:** Lifetime Dance Exposure Questionnaire for Professional Training
**MET:** metabolic equivalent
**PA:** physical activity
**REDCap:** Research Electronic Data Capture


*Edited by N Cahill; submitted 07.Apr.2025; peer-reviewed by L Krejcová; revised version received 11.Dec.2025; accepted 17.Dec.2025; published 27.Mar.2026.*

*Please cite as:*
*McCullough AK, Han KY(*
*The Lifetime Dance Exposure Questionnaire for Professional Training: Survey Development and Reliability Study*
*Interact J Med Res 2026;15:e75632*
*URL: https://www.i-jmr.org/2026/1/e75632*
*doi:10.2196/75632*


© Aston K McCullough, Kai Yi (Kaye) Han. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 27.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Understanding the Functional Needs of Patients With Cardiovascular Disease Regarding Nursing Robots Using a Kano Survey Scale: Cross-Sectional Study

XiuLi Wu[1], MSc; Aimei Kang[2], MEng

[1]Department of Nursing, Shandong Provincial Hospital Affiliated to Shandong First Medical University, No.324 Jingwu Road, Huaiyin District, Jinan, Shandong, China

[2]Department of Nursing, Wuhan Asia General Hospital Affiliated to Wuhan University of Science and Technology, WuHan, Hubei, China

**Corresponding Author:**
XiuLi Wu, MSc
Department of Nursing, Shandong Provincial Hospital Affiliated to Shandong First Medical University, No.324 Jingwu Road, Huaiyin District, Jinan, Shandong, China

## Abstract

**Background:** With population aging and the growing burden of chronic diseases, the number of patients with cardiovascular disease (CVD) in China continues to rise. Nursing robots have been recognized as one potential approach to help alleviate nursing workload. However, patient-centered functional needs with respect to such robots among patients with CVD remain insufficiently explored in China.

**Objective:** On the basis of the Kano model, this study aimed to analyze attitudes toward and demand priorities regarding nursing robot functions among patients with CVD. The findings may provide evidence to support function optimization and wider clinical application of nursing robots in China.

**Methods:** A cross-sectional study was conducted at Wuhan Asia Heart Hospital from July 2024 to June 2025. Convenience sampling was used to enroll 156 patients with CVD. Data were collected using a self-designed questionnaire, including demographic characteristics and a 15-item Kano scale for nursing robot functions. Data were analyzed using descriptive statistics, Kano classification, better-worse coefficients, and a 4-quadrant diagram.

**Results:** Among 15 functional items, accompanying care, consulting education, information management, and monitoring and warning were classified as must-be attributes; activity management, rehabilitation physiotherapy, drug management, security management, and logistics transfer were one-dimensional attributes; infection management, first aid skills, and nutritional care were attractive attributes; and sampling operation, personal hygiene management, and excretory care were indifferent attributes.

**Conclusions:** Our study reveals that patients with CVD generally exhibited high acceptance of nursing robots, but there were significant differences in functional requirements. When introducing nursing robots into clinical settings, priority should be given to essential attributes such as monitoring and warning, as well as educational consultation, along with expected attributes such as medication management. At the same time, attractive attributes such as infection management can be developed as value-added features. For nondifferentiated attributes such as personal hygiene management, selective configurations can be implemented based on clinical needs to enhance the clinical applicability and patient acceptance of nursing robots.

***(Interact J Med Res 2026;15:e92610)*** doi:10.2196/92610

**KEYWORDS**

artificial intelligence; AI; Kano model; demand analysis; robot; health services research

## Introduction

Worldwide, cardiovascular disease (CVD) has emerged as one of the primary risk factors adversely affecting public health [1]. According to a global statistical report by the American Heart Association, by 2021, there were over 19.91 million deaths attributed to CVD worldwide [2]. Over the past few decades, both the incidence and mortality rates of CVD in China have exhibited a significant upward trend. The number of patients with CVD in China alone amounts to 330 million [3]. Given the substantial population in China with risk factors for CVD, the fact that its incidence and mortality rates continue to rise poses significant challenges to the Chinese health care system [4].

Parallel to the growing CVD burden is the intensifying challenge of population aging and a shortage of professional nursing resources. Older adults account for a large proportion of patients with CVD and rely heavily on long-term, specialized care, yet the nursing workforce in China is underresourced and



XSL•FO
RenderX

overloaded, especially in cardiovascular departments, where nurses undertake repetitive, high-intensity tasks such as vital sign monitoring, medication guidance, and rehabilitation support [5-7]. In this context, nursing robots have emerged as a promising solution to alleviate workload, improve care efficiency, and supplement human resources in cardiovascular settings [8-11].

As a key component of intelligent health care, nursing robots have shown strong potential in supporting routine care and improving clinical delivery [12]. Current applications span health monitoring [13], daily living assistance [14,15], social companionship [16], and improvement of cognitive function [17]. Despite these advances, most research and functional designs focus on general older adult or rehabilitation populations, with limited consideration of the unique clinical needs of patients with CVD. Patients with CVD typically experience long disease courses, frequent exacerbations, polypharmacy, cardiac function–stratified rehabilitation, and elevated infection risk, creating distinct requirements for real-time monitoring, precise medication support, personalized rehabilitation guidance, and infection control [18]. Failure to address these specific needs may limit clinical translation and sustainable adoption of nursing robots in cardiovascular settings.

In recent years, research attention has shifted from technical development of robots to user needs and acceptance, with studies investigating perspectives of nursing managers, frontline nurses, and medical students [19-21]. As direct recipients of robot-assisted care, patients' views are critical for functional design and clinical adoption, yet their needs remain understudied.

The Kano model is a framework useful for categorizing user needs and ranking functional priorities and has been applied in some health care studies [18]. It can help classify needs into must-be, one-dimensional, attractive, and indifferent attributes. Previous studies have used the Kano model in medication management [22], pulmonary rehabilitation [23], psychological care [24], and nurses' needs for nursing robots [25]. However, few studies have used the Kano model to quantitatively explore the functional needs of patients with CVD regarding nursing robots.

Accordingly, this cross-sectional study used the Kano model to investigate the functional needs and priorities of patients with CVD regarding nursing robots. The findings may help bridge gaps between patient needs and robot function design and support the clinical translation of intelligent nursing technologies in cardiovascular care.

## Methods

### Study Design

A cross-sectional design was used to investigate functional needs regarding nursing robots among patients with CVD at a single time point.

### Research Setting

This study was conducted at Wuhan Asia Heart Hospital, the only tertiary specialized hospital in central China dedicated to the diagnosis and treatment of CVD. The hospital operates multiple departments, including cardiovascular medicine, cardiac surgery, and cardiovascular intensive care unit, with an annual patient volume of 128,000 for CVD. The patient population encompasses various conditions, such as atherosclerotic CVD, hypertensive CVD, and valvular heart disease. The distribution of patients is balanced in terms of age, educational background, disease duration, and severity, ensuring a strong representative sample that reflects the general needs of patients with CVD in central China.

### Participants

Convenience sampling was used to select patients with CVD who visited Wuhan Asia Heart Hospital from July 2024 to June 2025 as the primary study participants.

Inclusion criteria were (1) adults aged 18 years and older regardless of gender, (2) patients with a precise diagnosis of CVD (any type), and (3) voluntary participation in the study.

Exclusion criteria were (1) failure to provide valid informed consent or incomplete questionnaires, (2) individuals with significant cognitive impairment (a Mini-Mental State Examination score of <27), (3) those with decreased vision or hearing impairments making it impossible to read text or engage in conversation, (4) participants with critical acute conditions or those with end-stage disease, and (5) individuals who participated in a similar nursing robot needs survey within the previous 3 months.

### Sample Size Calculation

The minimum required sample size for the questionnaire was calculated using the following formula:

$$N=Z2\times p(1-p)2d2$$

A 95% CI ($Z$=1.96) and a 7.5% margin of error were applied. The minimum required sample size was 171. Considering an expected 10% invalid rate, 190 questionnaires were distributed.

### Research Tools

The survey tool comprised an online questionnaire designed with 2 sections. The first section was a general information survey that included 9 questions related to the participants' gender, marital status, age, educational level, type of CVD, duration of illness, previous experience with nursing robots, receipt of robot services, and methods of interaction with robots. The second part featured the Kano survey scale, which addressed functional needs regarding nursing robots. On the basis of a preliminary literature analysis, we identified and analyzed tasks in nursing that could potentially be replaced by robots, classifying them into 15 demand items [26,27] using card sorting methods (Table 1). In our developed Kano demand scale, each of the 15 included items was designed to assess participants' attitudes and preferences regarding specific functional areas of artificial intelligence (AI) nursing robots through both positive and negative questions. For instance, one question asked the following: "How do you feel if the nursing robot has/does not have a medication management function?" Each question encompassed 5 evaluation dimensions: "like it," "must-be," "neutral," "accept it," and "dislike." Consequently, participants' evaluations of a single item could yield 25 possible results (5



XSL•FO
RenderX

× 5), with each result corresponding to a specific attribute category (Table 2).

**Table .** Classification of functional areas of application of nursing robots.

| Category | Application scenario |
|---|---|
| Drug management | • Intravenous dispensing<br>• Drug distribution<br>• Medication reminder |
| Monitoring and warning | • Physiological indicator measurement and early warning<br>• Ward patrol<br>• Vital sign monitoring |
| Logistics transfer | • Transfer of patients<br>• Transfer of goods |
| Consulting education | • Medical information query<br>• Drug guidance<br>• Disease-related knowledge guidance<br>• Prevention of misinformation and education |
| Security management | • Prevention of falls and other accidental injuries<br>• Disaster prevention |
| Nutritional care | • Diet and nutrition guidance<br>• Feeding meals |
| Activity management | • Postural transition<br>• Bed-chair transfer<br>• Auxiliary walking |
| Excretory care | • Assistance with toileting |
| Personal hygiene management | • Organization and changing of bed linens<br>• Showering assistance<br>• Grooming (including brushing teeth, washing face, and shaving)<br>• Dressing and undressing |
| Accompanying care | • Interpersonal interaction<br>• Accompaniment for checkups<br>• Cognitive training |
| Infection management | • Ward cleaning, disinfection, and sterilization<br>• Wound disinfection and dressing |
| Rehabilitation physiotherapy | • Rehabilitation knowledge guidance<br>• Development and adjustment of the rehabilitation plan<br>• Assessment of rehabilitation effect |
| First aid skills | • Assisted CPR[a]<br>• Oxygen and sputum absorption |
| Information management | • Personal electronic medical information storage and record |
| Sampling operation | • Venous blood sampling<br>• Collecting nasal or pharyngeal swabs |

[a]CPR: cardiopulmonary resuscitation.



XSL•FO
RenderX

**Table .** Kano evaluation table.

| Positively worded items: how would you feel if the nursing robot had X function? | Negatively worded items: how would you feel if the nursing robot did not have X function? | | | | |
|---|---|---|---|---|---|
| | Like it | Must be | Neutral | Accept it | Dislike |
| Like it | Q | A[a] | A | A | O |
| Must be | R | I | I | I | M |
| Neutral | R | I | I | I | M |
| Accept it | R | I | I | I | M |
| Dislike | R | R | R | R | Q |

[a] Definition of categories in the Kano model. A = One-dimensional quality; M = Must-be quality; I = Indifferent quality; O = Attractive quality. One-dimensional quality means user satisfaction increases linearly with the fulfillment of requirements. Must-be quality is the basic requirement; dissatisfaction occurs if unfulfilled, while satisfaction cannot be greatly improved even if fulfilled. Attractive quality brings great satisfaction when realized and causes no dissatisfaction when absent. Indifferent quality has no influence on user experience. Reverse quality leads to lower satisfaction with higher implementation. Questionable quality represents contradictory and invalid responses.

## Reliability and Validity Test

Following the completion of the questionnaire design, we invited 5 experts—3 specializing in CVD nursing and 2 specializing in AI and communication technology—to evaluate our designed Kano demand scale. They assessed the relevance of the questions using Likert scales. The item-level content validity index for the Kano demand questionnaire ranged from 0.71 to 1.0, whereas the scale-level content validity index was 0.8. We assessed the effectiveness of the questionnaire using the Cronbach α coefficient, which was found to be 0.929. Additionally, internal consistency was evaluated through the retest coefficient, which yielded a value of 0.871. These results indicate that the effectiveness and reliability of the questionnaire were acceptable.

## Data Collection and Analysis

Participants were enrolled at Wuhan Asia Heart Hospital from July 2024 to June 2025. First, after completing the questionnaire design, the researchers created a survey using an online platform called Wenjuanxing (“Questionnaire Star”). Subsequently, the study objectives, content, risks, and benefits were explained to the participants, who voluntarily signed an electronic informed consent form. Finally, patients were guided to complete the questionnaire via mobile phones or tablets, with a completion time of 15 to 25 minutes. During this process, only the definitions of questionnaire items were explained, and no intervention was made on the participants’ choices. After the submission of the questionnaires, data collection began. During this phase, questionnaires exhibiting incomplete responses or evident logical contradictions were identified as invalid and excluded from subsequent analytical procedures.

SPSS (version 26.0; IBM Corp) was then used for data analysis. General data were described using descriptive statistics (numbers and percentage distributions). To comprehensively and accurately delineate the functional requirements regarding nursing robots among patients with CVD, 3 analytical approaches were used: the Kano model classification, the better-worse coefficient analysis for prioritization, and the construction of a Kano demand quadrant diagram.

First, cross-tabulation was performed on the positive and negative responses to each function in the questionnaire, categorized into 4 types (must attributes [M], one-dimensional attributes [O], attractive attributes [A], and indifferent attributes [I]). The frequency and proportion of each function being classified as different attributes were statistically analyzed, with the attribute type with the highest proportion serving as the final attribute classification for that function.

Second, we used the better-worse coefficient analysis method [23] to prioritize questionnaire items, thereby clarifying the impact level of specific functional domains on survey participants. This analytical method consists of “better” coefficients and “worse” coefficients, where the absolute values of the coefficients (ranging from 0 to 1) reflect satisfaction or dissatisfaction with the provision of services. Specifically, the closer the “better” value is to 1, the more satisfied patients are when the target attribute is provided. Conversely, the closer the absolute value of the “worse” coefficient is to 1, the higher the dissatisfaction level when the target attribute is not provided. The calculation formula for the coefficients is as follows:

$$\text{Better coefficient}=A+OA+O+M+I,\ \text{Worse coefficient}=O+MA+O+M+I$$

Finally, a 4-quadrant diagram was constructed with the “better” coefficient as the vertical axis and the absolute value of the “worse” coefficient as the horizontal axis. Specifically, “O” attributes located in the first quadrant (upper right region) exhibit both the “better” coefficient and the absolute value of the “worse” coefficient exceeding the mean, representing desirable needs that warrant greater attention and should be prioritized. “A” attributes with high “better” coefficients and low absolute values of the “worse” coefficient are situated in the second quadrant (upper left region), indicating desirable needs. “M” attributes in the fourth quadrant (lower right region) exhibit lower “better” coefficients and higher absolute values of the “worse” coefficient, representing essential attributes. “I” attributes in the third quadrant (lower left region) exhibit both lower “better” coefficients and lower absolute values of the “worse” coefficient, indicating nondifferentiated needs. These



XSL•FO
RenderX

needs are not significant to patients and may be omitted or do not require improvement [28].

### Ethical Considerations

The ethics committee of Wuhan University of Science and Technology (2024-096-02) approved this study, which adheres to the ethical standards outlined in the Declaration of Helsinki. Participation in this survey was voluntary and anonymous, and the participants provided informed consent. Participants could refuse or withdraw at any time, and the information collected was used solely for academic research purposes and not for commercial gain. All data were exclusively collected for this study, encrypted using a dual-password system (with passwords kept separately by 2 researchers), and stored on a hospital-specific research data server. The server undergoes regular security audits and backups. Data leakage, tampering, or use for other purposes are strictly prohibited. Upon completion of the survey, the data were archived in the hospital's research data platform for a period of 5 years, after which they will be uniformly destroyed in accordance with regulations.

## Results

### Basic Information of Survey Participants and Their Understanding of Nursing Robots

From July 2024 to June 2025, this study distributed 190 questionnaires, of which 6 (3.2%) were not completed due to privacy concerns. Of these 190 questionnaires, after excluding 28 (14.7%) that were invalid, a total of 156 (82.1%) valid questionnaires were collected. Although the final sample was slightly below the calculated target, post hoc evaluation indicated that the sample size was adequate for Kano classification and coefficient analysis, which rely on proportion comparisons rather than mean estimation.

The survey primarily focused on the Wuhan area of Hubei Province. The demographic characteristics of the study participants are summarized in Table 3. A total of 75.6% (118/156) of the participants were male, 45.5% (71/156) were aged over 60 years, 52.6% (82/156) had a bachelor's degree or higher, and 59% (92/156) had a disease duration of less than 10 years. Among the participants, 48.1% (75/156) had experience using nursing robots, 68.6% (107/156) had received nursing robot services, and 45.5% (71/156) preferred interactive methods such as images and videos.

### Demand Attribute Classification Analysis

Through the Kano model, we analyzed the requirements regarding nursing robots for patients with CVD across 15 functional areas. As shown in Table 4, based on the final survey results, we identified 3 items as "M" attributes: monitoring and warning, consulting education, and information management. Drug management and rehabilitation physiotherapy were categorized as "O" attributes. "A" attributes included 5 items: security management, logistics transfer, nutritional care, infection management, and first aid skills. The presence of these functional areas would make the nursing robots more acceptable to patients. The remaining 5 items were categorized as "I" attributes, which are not important to patients. These were excretory care, personal hygiene management, accompanying care, activity management, and sampling operation.



XSL•FO
RenderX

**Table .** Basic participant information (N=156).

| Variable | Participants, n (%) |
|---|---|
| Sex | |
| Male | 118 (75.6) |
| Female | 38 (24.4) |
| Marital status | |
| Married | 111 (71.2) |
| Unmarried | 45 (28.8) |
| Age (y) | |
| 18-25 | 37 (23.7) |
| 26-30 | 4 (2.6) |
| 31-40 | 36 (23.1) |
| 41-50 | 15 (9.6) |
| 51-60 | 3 (1.9) |
| >60 | 71 (45.5) |
| Educational background | |
| Primary school and lower | 19 (12.2) |
| Junior middle school | 31 (19.9) |
| High school or technical secondary school diploma | 24 (15.4) |
| Undergraduate or junior college | 73 (46.8) |
| Master's degree or higher | 9 (5.8) |
| Cardiovascular disease type | |
| Atherosclerotic cardiovascular disease | 44 (28.2) |
| Hypertensive cardiovascular disease | 19 (12.2) |
| Valvular heart disease | 29 (18.6) |
| Arrhythmia | 46 (29.5) |
| Cardiomyopathy | 11 (7.1) |
| Pericardial disease | 5 (3.2) |
| Other | 2 (1.3) |
| Duration of disease (y) | |
| <10 | 92 (59) |
| 10-20 | 36 (23.1) |
| >20 | 28 (17.9) |
| Previous use of nursing robots | |
| Yes | 75 (48.1) |
| No | 81 (51.9) |
| Receipt of robot services | |
| Yes | 107 (68.6) |
| No | 49 (31.4) |
| Methods of interaction with robots | |
| Text input | 47 (30.1) |
| Images and videos | 71 (45.5) |
| Speech recognition | 29 (18.6) |
| Other | 9 (5.8) |



XSL•FO
RenderX

**Table .** Classification of structural attributes of functional requirements for nursing robots among patients with cardiovascular disease (N=156)[a].

| Item | Attribute classification | | | | | | Final attribute classification |
|---|---|---|---|---|---|---|---|
| | M[b], n (%) | O, n (%) | A, n (%) | I, n (%) | R, n (%) | Q, n (%) | |
| Drug management | 2.56 | 35.9 | 26.28 | 17.95 | 16.03 | 1.28 | O |
| Monitoring and warning | 39.1 | 10.26 | 16.03 | 25 | 8.33 | 1.28 | M |
| Logistics transfer | 24.36 | 17.31 | 28.21 | 19.87 | 8.33 | 1.92 | A |
| Consulting education | 28.21 | 12.18 | 26.28 | 20.51 | 11.54 | 1.28 | M |
| Security management | 13.46 | 28.21 | 30.13 | 19.87 | 7.69 | 0.64 | A |
| Nutritional care | 5.77 | 12.82 | 50.64 | 21.15 | 8.33 | 1.28 | A |
| Activity management | 5.41 | 26.35 | 24.32 | 37.16 | 6.08 | 0.68 | I |
| Excretory care | 3.85 | 5.13 | 15.38 | 66.03 | 7.69 | 1.92 | I |
| Personal hygiene management | 3.85 | 7.05 | 14.1 | 67.31 | 7.05 | 0.64 | I |
| Accompanying care | 26.28 | 13.46 | 14.74 | 34.62 | 9.63 | 1.28 | I |
| Infection management | 4.49 | 6.41 | 53.85 | 27.56 | 7.05 | 0.64 | A |
| Rehabilitation physiotherapy | 3.21 | 35.9 | 19.87 | 30.77 | 8.97 | 1.28 | O |
| First aid skills | 5.13 | 11.54 | 47.44 | 24.36 | 8.97 | 2.56 | A |
| Information management | 37.82 | 9.62 | 24.36 | 17.95 | 8.97 | 1.28 | M |
| Sampling operation | 7.69 | 7.69 | 22.44 | 52.56 | 8.97 | 0.64 | I |

[a]Percentages are presented without subgroup n within cells. Each cell reflects joint results from two-dimensional Kano evaluation, and separate subgroup counts may cause misinterpretation. The overall sample size is reported in the main text.

[b] Definition of categories in the Kano model. A = One-dimensional quality; M = Must-be quality; I = Indifferent quality; O = Attractive quality; One-dimensional quality means user satisfaction increases linearly with the fulfillment of requirements. Must-be quality is the basic requirement; dissatisfaction occurs if unfulfilled, while satisfaction cannot be greatly improved even if fulfilled. Attractive quality brings great satisfaction when realized and causes no dissatisfaction when absent. Indifferent quality has no influence on user experience. ontradictory and invalid responses.

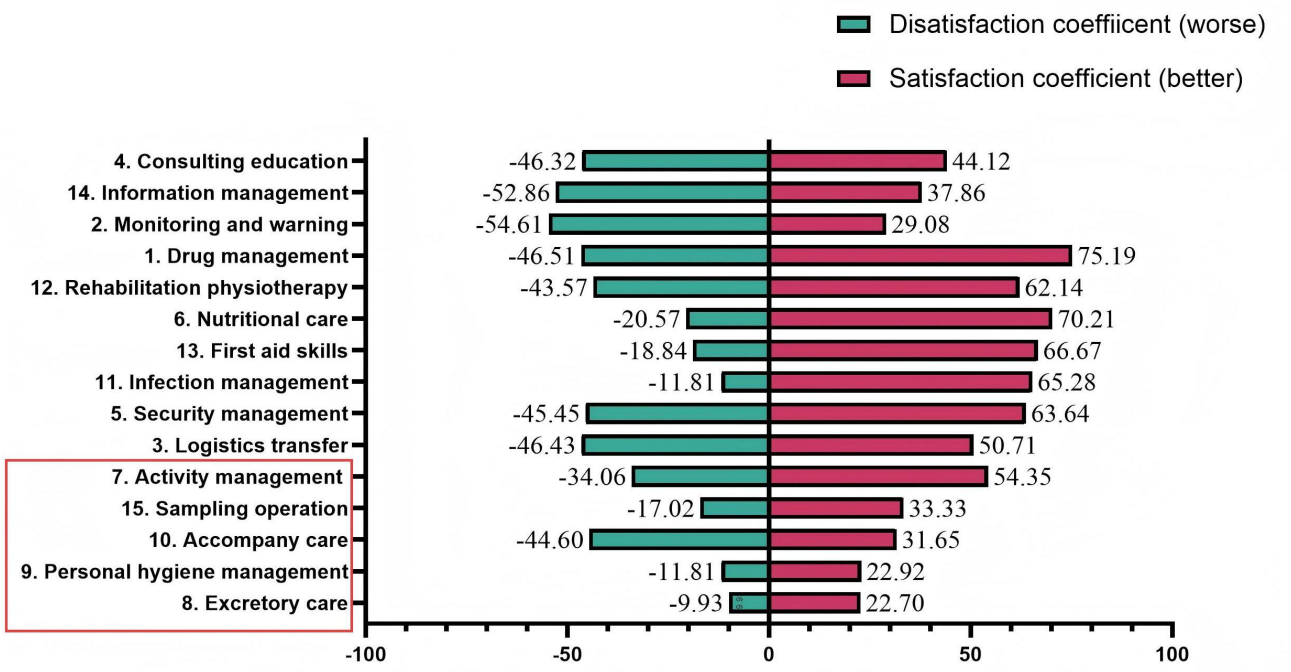
## Better-Worse Coefficient Analysis

To better visualize the functional differences across domains, we also created a comparative bar chart of “better” and “worse” coefficients for the functional requirements of patients with CVD regarding nursing robots. In Figure 1, among the top 10 functional areas, drug management, rehabilitation physiotherapy, nutritional care, first aid skills, infection management, security management, and logistics transfer had relatively high satisfaction index values ranging from 0.5 to 0.8. This indicates that, when the nursing robot possesses these functions, it will achieve high levels of satisfaction and acceptance. The other 3 functions (consulting education, information management, and monitoring and warning), although they had lower satisfaction index values, exhibited a high absolute dissatisfaction index value. This suggests that, if these requirements are not met, users will be highly dissatisfied regardless of how well other functions are performed, leading to a lack of increased favorability toward the product.



XSL•FO
RenderX

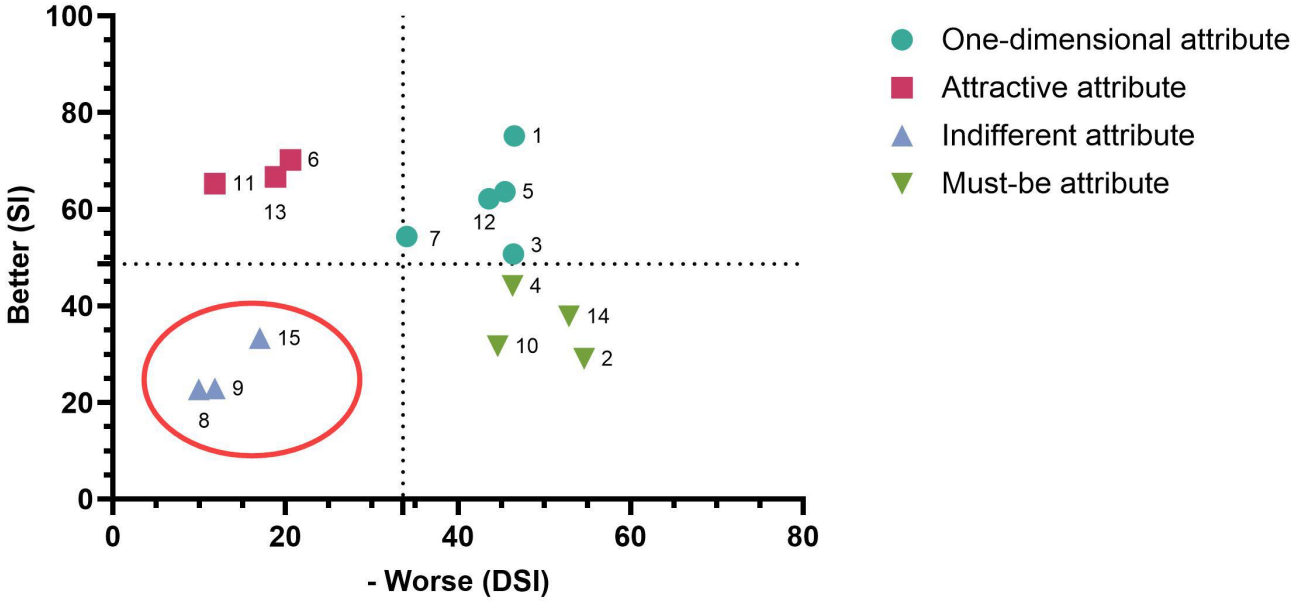
**Figure 1.** Satisfaction index and dissatisfaction index of functional requirements regarding nursing robots of patients with cardiovascular disease.




## Kano Demand Quadrant Diagram

We plotted a Kano demand quadrant diagram (Figure 2) using the better-worse coefficients. The results indicated that 4 items were categorized within the must-be attribute quadrant: monitoring and warning, consulting education, accompanying care, and information management. Five items were classified in the one-dimensional attribute quadrant: drug management, security management, logistics transfer, activity management, and rehabilitation physiotherapy. The attractive attribute quadrant included 3 items: nutritional care, infection management, and first aid skills. The remaining 3 items were placed in the indifferent attribute quadrant: excretory care, personal hygiene management, and sampling operation. Compared to the straightforward classification of needs presented in Figure 1, Figure 2 offers a more nuanced prioritization of each functional area, resulting in more precise outcomes. The findings illustrated in Figure 2 further corroborate that the functional areas of sampling operation, excretory care, and personal hygiene management are of relatively low priority.

**Figure 2.** Kano quadrant diagram of the demands of patients with cardiovascular disease regarding nursing robot functions. DSI: dissatisfaction index; SI: satisfaction index.






XSL•FO
RenderX

## Discussion

### Acceptance and Interaction Preference Characteristics Regarding Nursing Robots Among Patients With CVD

In this study, 68.6% (107/156) of participants with CVD had received nursing robot–assisted care, and 48.1% (75/156) reported prior exposure to the technology. This high familiarity likely stems from the 2-year clinical deployment of the Noah nursing robot in the study hospital's cardiology ward, which allowed patients to gradually build trust in this care model. This preexposure may have positively biased participants' attitudes toward nursing robots and, in turn, their perceptions of functional needs and Kano classification results. Our findings should therefore be generalized with caution to settings without routine nursing robot use and further validated in larger, multicenter cohorts across different hospital tiers and regions.

For interaction preferences, 45.5% (71/156) of patients preferred visual or video-based interaction, 30.1% (47/156) favored text input, and only 18.6% (29/156) chose voice recognition. This pattern is closely tied to the 45.5% (71/156) of older adult patients in our cohort: visual and video formats deliver medication guidance and rehabilitation instructions more intuitively for older adults, whereas many patients also voiced concerns about the accuracy of voice recognition in busy wards and associated privacy risks, leading to a preference for traceable, visual interaction.

Consistent with previous work [29], voice interaction still has clear potential for efficient human-machine communication. Our findings suggest that nursing robot design should prioritize multimodal interaction, with targeted optimization of voice recognition for older adult patients' dialects plus accessible interaction record functions to strengthen privacy and security for users.

### Structural Analysis of the Functional Needs of Patients With CVD Regarding Nursing Robots Based on the Kano Model

Using the Kano model, we categorized 15 core nursing robot functions into 4 attributes: must be (M), one-dimensional (O), attractive (A), and indifferent (I). This classification clarifies the clinical positioning and optimization priorities of each function, providing evidence-based guidance for the design and clinical deployment of nursing robots for CVD care.

"M" attributes included monitoring and early warning, information management, and consultation and education, all of which align with the core needs of long-term monitoring and self-management for patients with CVD. Notably, the monitoring and early warning function had an absolute "worse" coefficient of 54.6, meaning that more than half of the patients would report reduced satisfaction if this function were absent. This is consistent with the chronic, fluctuating nature of CVD, where real-time monitoring and timely warning are critical to prevent adverse events [30]. These core basic functions should be prioritized in robot design as they form the foundation of clinical acceptance and align with the shift toward proactive, precision smart nursing.

Medication management and rehabilitation physiotherapy were classified as "O" attributes, with "better" coefficients over 60 for both, showing that optimizing these functions directly improves patient satisfaction. Patients with CVD typically take an average of 3.2 concurrent medications, with complex regimens carrying a high risk of dosing errors and nonadherence [31], explaining the strong demand for integrated medication reminders, dose verification, and supply alerts. For rehabilitation, patients preferred personalized plans stratified by New York Heart Association cardiac function class over standardized programs, reflecting the need for personalized, rather than one-size-fits-all, smart nursing support.

Infection management, nutritional care, and first aid skills were identified as "A" attributes. While not mandatory for patients, these functions drive significant improvements in satisfaction. The infection management function had a "better" coefficient of 65.3, which aligns with the higher infection risk in immunocompromised patients with CVD [32]. Personalized nutrition plans tailored to lipid and glucose levels also fit the model of precision nutritional support, making these functions useful value-added features for high-acuity settings such as cardiovascular intensive care units.

Personal hygiene management, excretion care, and sampling operations were classified as "I" attributes, with minimal patient demand for robot delivery of these services. This low demand is likely driven by 2 key factors: privacy concerns regarding highly personal care tasks and patient preference for functions that directly improve self-care safely rather than high-risk, specialized operations. This preference for robot involvement in noninvasive, nontechnical care aligns with findings of previous studies [17,33,34]. Routine large-scale deployment of invasive functions such as blood sampling is therefore not recommended; these may only be selectively used in specific settings such as isolation wards. Optimizing privacy protection features will also be key to addressing patient trust concerns around these functions.

Overall, the clinical deployment of nursing robots for CVD care should be guided by patient needs following a hierarchical optimization framework: prioritizing essential must-be functions, optimizing one-dimensional functions to improve care quality, adding attractive functions as value-added features, and limiting indifferent functions to scenario-specific use. This framework aligns robot design and technical development with clinical needs, supporting the practical advancement of smart nursing.

### Implications for Clinical Practice and Future Research

Our findings have direct implications for the clinical deployment of nursing robots in Wuhan and beyond, as well as for the advancement of patient-centered intelligent cardiovascular nursing. Locally, the hierarchical priority structure we identified can be used to refine existing nursing robot programs in Wuhan hospitals, with prioritized ward deployment of high-demand functions such as monitoring and early warning, medication management, and nutritional support to align services with patient needs.

For routine nursing practice, integrating robots into high-demand processes such as medication monitoring can standardize care,



XSL•FO
RenderX

reduce human error, and improve continuity of care. Rather than large-scale changes to existing nursing workflows, we recommend a phased approach: piloting high-priority functions, evaluating safety and acceptability, and then gradually expanding successful modules.

These findings also highlight key directions for future research. Larger multicenter studies are needed to validate the generalizability of these demand patterns, whereas subgroup analyses can clarify heterogeneous needs across patient groups. Longitudinal studies evaluating the safety, usability, and clinical outcomes of robot-assisted care will also be critical to support evidence-based translation into practice.

Ultimately, this study addresses a key gap between nursing robot technical development and real-world clinical needs by centering patient perspectives in function design. The proposed hierarchical framework supports a shift from generalized robot deployment to precision, patient-centered intelligent care, which has the potential to reduce nursing workload, improve care standardization, and enhance patient safety and satisfaction in cardiovascular care.

### Limitations

This study has several limitations. First, the sample was limited to 1 hospital in Wuhan using convenience sampling, which may reduce generalizability. Second, subgroup analyses by age, educational level, or disease severity were not performed, although age stratification was explored and showed limited between-group differences in this cohort. Third, this was a cross-sectional self-reported survey, which may be subject to response bias. Future studies should use larger, multicenter samples; perform stratified analyses; and include longitudinal follow-up to strengthen the evidence base.

### Summary

The presence of robots and the advent of more advanced robotics technologies have led to an increase in studies assessing user acceptance of robots and needs analyses in the past 2 years. Wichmann et al [35] showed that perceived usefulness, attitude, and trust significantly and positively predict willingness to use AI and use behaviors in short- and long-term treatment. In our view, clarifying the needs of patients can effectively improve users' acceptance of AI and may become a shortcut to achieving intelligent medical care in China. Our study explores the differences in prioritizing needs regarding nursing robots from the perspective of patients with CVD. This paper suggests that researchers should emphasize understanding the patient experience, which is crucial for advancing various fields, including counseling, education, companion care, and more. We hope that our research can help nursing robot developers design targeted functional services for patients with CVD, promoting the continuous development and application of nursing robotics as well as innovation and progress in the medical field.

## Acknowledgments

First, the authors would like to thank the nurses and head nurses at Wuhan Asia Heart Hospital for their support in this study. Second, they sincerely thank the editors, statisticians, and anonymous reviewers for their insightful comments, which greatly improved the paper. The authors declare the use of generative artificial intelligence (GenAI) in the research and writing process. According to the Generative Artificial Intelligence Delegation Taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: preparation of press releases and outreach materials. The GenAI tool used was Doubao (ByteDance). Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes. In addition, the authors used Doubao to generate the table of contents image for this paper.

## Funding

The authors declared no financial support was received for this work.

## Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

## Conflicts of Interest

None declared.

## References

1. Cardiovascular diseases. World Health Organization. URL: https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1 [accessed 2026-06-28]
2. Martin SS, Aday AW, Almarzooq ZI, et al. 2024 Heart Disease and Stroke Statistics: a report of US and global data from the American Heart Association. Circulation 2024 Feb 20;149(8):e347-e913. [doi: 10.1161/CIR.0000000000001209] [Medline: 38264914]
3. Writing Committee of the Report on Cardiovascular Health and Diseases in China. Report on cardiovascular health and diseases in China 2022: an updated summary. Biomed Environ Sci 2023 Aug 20;36(8):669-701. [doi: 10.3967/bes2023.106] [Medline: 37711081]



XSL•FO
RenderX

4. Zheng C, Wang X, Gu R, et al. The burden of cardiovascular disease in China: incidence estimates from national surveillance (2023). J Am Coll Cardiol 2025 Oct 8:S0735-1097(25)07615-6. [doi: 10.1016/j.jacc.2025.08.060] [Medline: 41263755]
5. Bao J, Zhou L, Liu G, et al. Current state of care for the elderly in China in the context of an aging population. Biosci Trends 2022 May 17;16(2):107-118. [doi: 10.5582/bst.2022.01068] [Medline: 35431289]
6. Ni W, Xia M, Jing M, Zhu S, Li L. The relationship between professional quality of life and work environment among ICU nurses in Chinese: a cross-sectional study. Front Public Health 2023;11:1104853. [doi: 10.3389/fpubh.2023.1104853] [Medline: 37213646]
7. Tamata AT, Mohammadnezhad M. A systematic review study on the factors affecting shortage of nursing workforce in the hospitals. Nurs Open 2023 Mar;10(3):1247-1257. [doi: 10.1002/nop2.1434] [Medline: 36303066]
8. Douthit BJ, Richesson RL, Kim H, Cary MP. How artificial intelligence is transforming the future of nursing: thinking about care in a new light. Am Nurs J 2020;15(9):100 [FREE Full text]
9. Jaltotage B, Lu J, Dwivedi G. Use of artificial intelligence including multimodal systems to improve the management of cardiovascular disease. Can J Cardiol 2024 Oct;40(10):1804-1812. [doi: 10.1016/j.cjca.2024.07.014] [Medline: 39038650]
10. Khera R, Oikonomou EK, Nadkarni GN, et al. Transforming cardiovascular care with artificial intelligence: from discovery to practice: JACC state-of-the-art review. J Am Coll Cardiol 2024 Jul 2;84(1):97-114. [doi: 10.1016/j.jacc.2024.05.003] [Medline: 38925729]
11. Zhang S, Liu L, Peng T, Ding S. Patterns and trends in global nursing robotics research: a bibliometric study. J Nurs Manag 2025;2025:7853870. [doi: 10.1155/jonm/7853870] [Medline: 40223900]
12. Ng ZQ, Ling LY, Chew HS, Lau Y. The role of artificial intelligence in enhancing clinical nursing care: a scoping review. J Nurs Manag 2022 Nov;30(8):3654-3674. [doi: 10.1111/jonm.13425] [Medline: 34272911]
13. Tian Y, Chen C, Sagoe-Crentsil K, Zhang J, Duan W. Intelligent robotic systems for structural health monitoring: applications and future trends. Autom Constr 2022 Jul;139:104273. [doi: 10.1016/j.autcon.2022.104273]
14. Bovbel P, Nejat G. Casper: an assistive kitchen robot to promote aging in place. J Med Devices 2014 Sep;8(3):030945. [doi: 10.1115/1.4027113]
15. Ko JB, Kong YK, Choi KH, et al. Comparison of the physical care burden on formal caregivers between manual human care using a paper diaper and robot-aided care in excretion care. Int J Environ Res Public Health 2023 Jan 10;20(2):1281. [doi: 10.3390/ijerph20021281] [Medline: 36674037]
16. Rasouli S, Gupta G, Nilsen E, Dautenhahn K. Potential applications of social robots in robot-assisted interventions for social anxiety. Int J Soc Robot 2022;14(5):1-32. [doi: 10.1007/s12369-021-00851-0] [Medline: 35096198]
17. Esposito R, Fiorini L, Limosani R, et al. Supporting active and healthy aging with advanced robotics integrated in smart environment. In: Morsi YS, Shukla A, Rathore CP, editors. Optimizing Assistive Technologies for Aging Populations: IGI Global Scientific Publishing; 2016:46-77. [doi: 10.4018/978-1-4666-9530-6]
18. Liu X, He X, Wang M, Shen H. What influences patients' continuance intention to use AI-powered service robots at hospitals? The role of individual characteristics. Technol Soc 2022;70(6):101996. [doi: 10.1016/j.techsoc.2022.101996]
19. Ergin E, Karaarslan D, Şahan S, Çınar Yücel Ş. Artificial intelligence and robot nurses: from nurse managers' perspective: a descriptive cross-sectional study. J Nurs Manag 2022 Nov;30(8):3853-3862. [doi: 10.1111/jonm.13646] [Medline: 35474366]
20. Lee JY, Song YA, Jung JY, et al. Nurses' needs for care robots in integrated nursing care services. J Adv Nurs 2018 May 13. [doi: 10.1111/jan.13711] [Medline: 29754395]
21. Łukasik S, Tobis S, Kropińska S, Suwalska A. Role of assistive robots in the care of older people: survey study among medical and nursing students. J Med Internet Res 2020 Aug 12;22(8):e18003. [doi: 10.2196/18003] [Medline: 32784187]
22. Mao J, Xie L, Zhao Q, et al. Demand analysis of an intelligent medication administration system for older adults with chronic diseases based on the Kano model. Int J Nurs Sci 2021;9(1):63-70. [doi: 10.1016/j.ijnss.2021.12.012] [Medline: 35079606]
23. Yao X, Li J, He J, et al. A Kano model-based demand analysis and perceived barriers of pulmonary rehabilitation interventions for patients with chronic obstructive pulmonary disease in China. PLoS One 2023;18(12):e0290828. [doi: 10.1371/journal.pone.0290828] [Medline: 38109304]
24. Wang Z, Tang X, Li L, et al. Spiritual care needs and their attributes among Chinese inpatients with advanced breast cancer based on the Kano model: a descriptive cross-sectional study. BMC Palliat Care 2024 Feb 22;23(1):50. [doi: 10.1186/s12904-024-01377-8] [Medline: 38388378]
25. Gu Y, Shi T, Guo S, et al. A survey of clinical nurses' needs for nursing robot applications based on the Kano model. J Nurs 2023;38(11):53-58. [doi: 10.3870/j.issn.1001-4152.2023.11.053]
26. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health 2021 Sep;3(9):e599-e611. [doi: 10.1016/S2589-7500(21)00132-1] [Medline: 34446266]
27. Kang A, Wu X. Assessing visitor expectations of AI nursing robots in hospital settings: cross-sectional study using the Kano model. JMIR Nurs 2024 Nov 27;7:e59442. [doi: 10.2196/59442] [Medline: 39602413]
28. Wang T, Zhao Y, Zhang X, Xie Y, Pang LL. Design of walking aids for the elderly based on the Kano-AHP-FEC method. Sci Rep 2025;15(1):2663. [doi: 10.1038/s41598-025-85540-y]




XSL•FO
RenderX

29. Marge M, Espy-Wilson C, Ward NG, et al. Spoken language interaction with robots: recommendations for future research. Comput Speech Lang 2022 Jan;71:101255. [doi: 10.1016/j.csl.2021.101255]
30. Ullah M, Hamayun S, Wahab A, et al. Smart technologies used as smart tools in the management of cardiovascular disease and their future perspective. Curr Probl Cardiol 2023 Nov;48(11):101922. [doi: 10.1016/j.cpcardiol.2023.101922] [Medline: 37437703]
31. Brajković A, Bosnar L, Nascimento MM, et al. Healthcare utilisation and clinical outcomes in older cardiovascular patients receiving comprehensive medication management services: a nonrandomised clinical study. Int J Environ Res Public Health 2022 Feb 27;19(5):2781. [doi: 10.3390/ijerph19052781] [Medline: 35270472]
32. Sipilä PN, Lindbohm JV, Batty GD, et al. Severe infection and risk of cardiovascular disease: a multicohort study. Circulation 2023 May 23;147(21):1582-1593. [doi: 10.1161/CIRCULATIONAHA.122.061183] [Medline: 36971007]
33. Komagata T, Sasaki Y, Yoshimitsu K, et al. Bridging between clinical nursing practice and robotics: which nursing tasks can be delegated to robots? Adv Robot 2025;39(23):1549-1557. [doi: 10.1080/01691864.2025.2587835]
34. Zeng Z, Chen PJ, Lew AA. From high-touch to high-tech: COVID-19 drives robotics adoption. Tour Geogr 2020 May;22(3):724-734. [doi: 10.1080/14616688.2020.1762118]
35. Wichmann J, Gesk TS, Leyer M. Acceptance of AI in health care for short- and long-term treatments: pilot development study of an integrated theoretical model. JMIR Form Res 2024 Jul 18;8:e48600. [doi: 10.2196/48600] [Medline: 39024565]

## Abbreviations

**AI:** artificial intelligence
**CVD:** cardiovascular disease

*Edited by M Balcarras; submitted 01.Feb.2026; peer-reviewed by IA Eide, S Hao; revised version received 06.May.2026; accepted 25.May.2026; published 25.Aug.2026.*

*Please cite as:*
*Wu X, Kang A*
*Understanding the Functional Needs of Patients With Cardiovascular Disease Regarding Nursing Robots Using a Kano Survey Scale: Cross-Sectional Study*
*Interact J Med Res 2026;15:e92610*
*URL: https://www.i-jmr.org/2026/1/e92610*
*doi:10.2196/92610*

© Aimei Kang, XiuLi Wu. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 25.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Identification of the Core Competencies Required in Endodontics for Undergraduate Students in Syrian Dental Schools by Using a Modified Delphi Technique: Prospective Exploratory Survey Study

Muhammad Salameh[1,2*], MSc, DDS, PhD; Mayssoon Dashash[1,3*], MSc, DDS, PhD; Issam Jamous[1,4*], MSc, DDS, PhD

[1]Medical Education Program, Syrian Virtual University, Damascus, Dimashq, Syrian Arab Republic

[2]Department of Endodontics, Faculty of Dentistry, Al Andalus University, Tartus, Syrian Arab Republic

[3]Department of Pediatric Dentistry, Faculty of Dentistry, Damascus University, Damascus, Dimashq, Syrian Arab Republic

[4]Department of Fixed Prosthodontics, Faculty of Dentistry, Damascus University, Damascus, Dimashq, Syrian Arab Republic

[*]all authors contributed equally

**Corresponding Author:**
Muhammad Salameh, MSc, DDS, PhD
Medical Education Program
Syrian Virtual University
Medical Education Program
Damascus, Dimashq
Syrian Arab Republic
Phone: 963 944273017
Email: drsalameh970@gmail.com

## Abstract

**Background:** There is a worldwide movement toward competency-based medical education to equip dental students with essential competencies required to meet health care needs. In Syria, dental faculties currently lack a formal competency-based curriculum for endodontics at the undergraduate level. Moreover, the quality of root canal treatment performed by general dentists is frequently described as inadequate or substandard.

**Objective:** This study aimed to develop a national consensus on the required competencies for undergraduate endodontics in Syria in order to establish a foundation for a standardized national curriculum, which can guide educators in adopting best practices in both dental education and clinical endodontics.

**Methods:** This study was conducted at Syrian Virtual University between April and June 2025. A modified Delphi technique was used to determine endodontic competencies. Initially, a group of 5 Syrian endodontic consultants identified preliminary competencies. In the first round, 53 experts evaluated these competencies by using a 5-point Likert scale. Based on these results, a second round was conducted with 38 experts. Competencies with a weighted average above 4.20 were considered essential. Data analysis was performed using IBM SPSS package 27, and survey reliability was measured by Cronbach α.

**Results:** Following the final Delphi round, a set of 31 competencies was established, comprising 9 knowledge, 13 skills, and 9 attitudes competencies. Cronbach α was more than 0.9 in the first and second round. The standard deviation across all questionnaires was low (≤0.85). The standard error was also minimal (≤0.12).

**Conclusions:** This study identified a set of core endodontic competencies for the undergraduate level in Syria. These competencies are intended to support students in acquiring the required knowledge, skills, and attitudes, and assisting policymakers in implementing competency-based medical education within Syria and similar contexts.

***(Interact J Med Res 2026;15:e83799)*** doi:10.2196/83799

**KEYWORDS**

endodontics; Syria; undergraduate curriculum; competency-based medical education; modified Delphi technique; Delphi technique; consultation meeting; root canal treatment



XSL•FO
RenderX

## Introduction

General dentists frequently serve as the primary providers of endodontic treatment for patients requiring such care [1,2]. However, cross-sectional studies often report the quality of root canal treatment as inadequate or below standard [3-7]. This may be attributed to the complexity of root canal procedures or the graduation of practitioners with limited experience [8,9].

The Commission of Dental Accreditation, the Association of Dental Education in Europe, and the European Society of Endodontology (ESE) have all recommended adopting competency-based medical education (CBME) within a comprehensive clinical care environment [8,10,11].

According to the International CBME collaborators, CBME is "an outcome-based approach to the design, implementation, assessment, and evaluation of medical education programs using an organized framework of competencies" [12]. The goal of CBME is to produce graduates who can effectively incorporate knowledge, skills, and attitudes and apply them clinically to deliver the best patient care [13]. CBME emphasizes outcomes and practice-based learning while considering societal and patient needs by tailoring competency frameworks accordingly [14]. As a result, several countries have adopted CBME into undergraduate medical education [15,16] although its application remains in early stages [16].

In Syria, medical education follows a traditional, time-based approach that emphasizes teacher-centered learning, hospital-based training, and opportunistic learning methods [17]. To our knowledge, there is no existing literature detailing the design and implementation of curricula in Syrian medical schools. Typically, curriculum development and review are undertaken by a small group of academic members without formal consensus from all faculty members before final approval by the faculty and university councils [18].

Competency-based education is a fundamental part of the academic curriculum, as it measures and verifies the attainment of the intended program learning outcomes [19]. Therefore, the initial step in implementing a CBME is to identify the required competencies [20]. This method offers medical students focused and well-structured training in endodontics, ensuring they acquire the necessary knowledge, skills, and attitudes to provide effective patient care. Within this framework, this study aims to establish a national consensus on the core competencies required for undergraduate endodontics curricula in Syria.

CBME considers competencies as the key outcomes that should guide curriculum development at every stage, including implementation, assessment, and evaluation [21-23]. To embrace CBME and adequately prepare medical students for clinical practice, educators usually rely on an organized national or international competency framework that describes the abilities physicians need to meet patients' and societal demands [24].

The Delphi technique, a widely used consensus method for identifying health professional competencies, has been applied independently [25] or alongside focus groups [26] and consultation meetings [27]. In this study, a competency framework for undergraduate endodontics was developed using a combination of the Delphi technique and consultation meetings to identify essential competencies. Therefore, this study aims to develop a national consensus on essential undergraduate endodontic competencies in Syria by using a modified Delphi method.

## Methods

### Overview

This exploratory qualitative study was undertaken at the Syrian Virtual University from April to June 2025. The modified Delphi technique was used to reach a consensus on the competencies for the undergraduate endodontics curriculum.

The consultation group included 5 Syrian members holding a PhD in endodontics and working in academic roles at various universities. Additionally, experts were either Syrians or held equivalent qualifications, possessed at least a specialist certificate (Syrian board) in endodontics, and they were currently actively involved in clinical, academic, and/or administrative practice inside or outside Syria.

The principal researcher (MS) developed a preliminary list of endodontic competencies based on a literature review [28-30]. Individuals' consultation meetings were then held via telephone or social media platforms, lasting at least 15 minutes each, and conducted at least twice with each consultant. These meetings reviewed the initial list of competencies, allowing for modifications and additions. The consultation meetings produced a final list of 54 competencies divided into 19 knowledge, 21 skills, and 14 attitude items. A structured questionnaire was developed using Google Forms, consisting of six sections: (1) an introductory letter explaining the study's objectives, confidentiality assurance, participation instructions, and consent acknowledgement; (2) demographic data, including name, age, gender, nationality, workplace, years of experience, academic qualifications, and practice type; (3,4,5) the initial lists of knowledge, skills, and attitudes competencies in the third, fourth, and fifth sections, respectively; and (6) a free-text section for respondents to suggest additional competencies (Multimedia Appendix 1).

The Delphi methodology was conducted in 2 rounds to identify the essential endodontic competencies for undergraduate dental students in Syrian dental schools. The principal researcher MS invited 98 endodontists who met the expert panel criteria to participate, with 73 accepting the invitation. To ensure a national consensus, these experts represented dental schools from both public and private universities across Syria. Participants rated each competency on a 5-point Likert scale reflecting its importance: 1 (not important at all), 2 (not important), 3 (unsure), 4 (important), and 5 (very important). Consensus was defined as over 80% agreement or a mean score above 4.20 [25,31].

Strict deadlines were set to maintain study momentum while allowing participant flexibility with accountability to ensure a smooth and efficient progression through its phases [32]. The first Delphi round lasted 3 weeks, starting on May 2, 2025, with reminders sent to nonrespondents on the 5th, 10th, and 15th. Competencies scoring above 4.20 in the first round, along with



XSL•FO
RenderX

those suggested through free-text comments, were included in the second round. The second round had a 2-week deadline, with reminders on days 5 and 10 for those who had not responded.

Data analysis was performed using the SPSS package (version 27; IBM Corp). The survey's reliability was assessed using Cronbach α, and descriptive statistics, including the mean, standard deviation, and standard error, were calculated.

### Ethical Considerations

Ethical approval was obtained from the research committee of the Syrian Virtual University (455/0, dated April 27, 2025). Informed consent was obtained from all participating consultants and experts before the study.

## Results

The characteristics of the participants in both the first and second Delphi rounds are shown in Table 1. Following the consultation meetings, a preliminary list of 54 competencies was established, comprising 19 (35.2%) knowledge competencies, 21 (38.8%) skills competencies, and 14 (26%) attitudes competencies (Multimedia Appendix 1). In the first Delphi round, 53 out of the 73 participants completed the survey, yielding a response rate of 72.6%.

**Table 1.** Characteristics of the respondents in the two Delphi rounds.

| Characteristics | Round 1, n (%) | Round 2, n (%) |
|---|---|---|
| Sample | | |
| Participants | 73 (100) | 73 (100) |
| Respondents | 53 (72.6) | 38 (52) |
| Gender | | |
| Male | 40 (75.5) | 27 (71) |
| Female | 13 (24.5) | 11 (29) |
| Qualifications | | |
| PhD | 38 (71.7) | 27 (71) |
| Master | 13 (24.5) | 9 (23.7) |
| Other | 2 (3.8) | 2 (5.3) |
| Years of experience | | |
| ≤ 5 y | 6 (11.3) | 2 (5.3) |
| 6-10 y | 13 (24.5) | 9 (23.7) |
| 11-15 y | 10 (18.8) | 7 (18.4) |
| 16-20 y | 7 (13.2) | 9 (23.7) |
| 21-25 y | 10 (18.9) | 4 (10.5) |
| 25 y | 7 (13.2) | 7 (18.4) |
| Type of work | | |
| Academic only | 14 (26.4) | 14 (36.8) |
| Clinical only | 11 (20.7) | 9 (23.7) |
| Academic & clinical | 22 (41.5) | 10 (26.3) |
| Academic, clinical, & administrative | 6 (11.3) | 5 (13.2) |

In the first Delphi round, 34 competencies were approved, consisting of 10 knowledge, 15 skills, and 9 attitudes. Additionally, participants suggested 4 new competencies, 1 knowledge, and 3 attitudes, bringing the total to 38 for the second round (Multimedia Appendix 2). These newly suggested competencies were evaluated by participants using the same criteria as in the first round. In the second Delphi round, 38 out of 73 experts completed the survey, resulting in a 52% response rate. Table 1 presents the characteristics of the second-round participants. Of the 38 competencies, 31 (81.5%) received mean ratings above 4.20. These were distributed across the domains as follows: 9 (29%) knowledge, 13 (42%) skills, and 9 (29%) attitude (Table 2). In this round, competencies were ranked based on experts' weighted responses.



XSL•FO
RenderX

**Table 2.** Final competencies in endodontics for the undergraduate level.

| Competency | Mean | SE | SD |
|---|---|---|---|
| Knowledge | | | |
| Diagnostic investigations, 2D and 3D radiography | 4.68 | .083 | .518 |
| Dental materials, biomaterials, and adjunct therapies applied to the management of endodontics | 4.53 | .080 | .499 |
| Oral and dental diseases related to endodontics | 4.42 | .079 | .494 |
| Basic sciences and their relationship to endodontics | 4.39 | .078 | .489 |
| Pharmacology and therapeutics as applied to the management of dental patients | 4.39 | .094 | .587 |
| Principles of general medicine and surgery applied to the management of dental patients (including endodontics) | 4.24 | .093 | .582 |
| Principles of optical magnification | 4.24 | .107 | .666 |
| Principles of management of immature teeth | 4.21 | .098 | .614 |
| Treatment options for a postendodontic problem | 4.21 | .111 | .694 |
| Skill | | | |
| Management of endodontic emergencies | 4.58 | .079 | .494 |
| Reliable and appropriate isolation | 4.53 | .080 | .499 |
| Use various intraoral anesthesia techniques and pain control | 4.50 | .080 | .500 |
| Establishing a reliable root canal irrigation protocol | 4.50 | .088 | .550 |
| Diagnose and differentiate odontogenic pain or lesions | 4.47 | .080 | .499 |
| Reaching a diagnosis and identifying possible differential diagnoses, including their etiology | 4.42 | .125 | .782 |
| Performing high-quality endodontic treatments on extracted or simulated (Acrylic Blocks) teeth of various types | 4.42 | .114 | .712 |
| Perform high-quality clinical endodontic treatment using conventional methods for easy or moderately difficult teeth (according to the American Association of Endodontics classification) | 4.42 | .087 | .544 |
| Develop a treatment plan and communicate this to the patient | 4.39 | .101 | .630 |
| Restoration of endodontically treated teeth, including root canal posts | 4.37 | .100 | .625 |
| Conducting a comprehensive clinical examination of a patient presenting with an endodontic-related problem | 4.34 | .123 | .770 |
| Assess the case difficulty and request a consultation or refer the patient | 4.34 | .123 | .770 |
| Conducting a detailed general and dental history | 4.24 | .107 | .666 |
| Attitudes | | | |
| Respect patient privacy and confidentiality | 4.66 | .076 | .474 |
| Self-efficacy | 4.42 | .087 | .544 |
| Time and priority management | 4.42 | .087 | .544 |
| Commitment to lifelong learning | 4.37 | .100 | .625 |
| Compliance with local regulations regarding infection control, radiation protection, record keeping, and documentation | 4.37 | .100 | .625 |
| Handling medical waste | 4.37 | .093 | .581 |
| Communicate verbally and in writing with dental and medical colleagues | 4.34 | .112 | .699 |
| Communicate effectively with the patient or their family | 4.34 | .099 | .619 |
| Informed consent | 4.34 | .099 | .619 |

Based on the results of the second Delphi round, the final list of competencies for the undergraduate endodontics curriculum was finalized (Table 2).

The standard deviation for all questionnaire items was low, not exceeding 0.85, and the standard error was also minimal, remaining below 0.12. These values indicate a close alignment between experts' individual ratings and the overall mean. Additionally, Cronbach α exceeded 0.9 for both the first and second rounds, demonstrating a high reliability level of internal consistency reliability (Table 3).



XSL•FO
RenderX

**Table 3.** Cronbach α results in both Delphi rounds.

| Domain of competency | Total | Attitude | Skill | Knowledge |
|---|---|---|---|---|
| First round | 0.901 | 0.783 | 0.828 | 0.863 |
| Second round | 0.909 | 0.832 | 0.845 | 0.749 |

# Discussion

## Principal Findings

The findings of this study support a national consensus among experts regarding the core competencies required in endodontics at the undergraduate level in Syria. A set of 31 competencies was established, comprising 9 knowledge, 13 skills, and 9 attitudes.

There is no universally accepted definition of endodontic competency for the new general dentist [33]. Moreover, there are currently no national or international guidelines for undergraduate endodontic education specifying the essential materials and equipment that students should be competent in using [34]. This study was conducted to address that gap.

Developing a national or international framework for undergraduate endodontic education can help educators reach consensus and implement best practices in education and endodontics [29]. Clearly defining the required materials and equipment is crucial for standardized practical training [34]. The primary goal of undergraduate endodontic education is to produce competent clinicians who understand the basic science of endodontics and recognize limitations, ensuring safe general dental practice. Clinical training should also emphasize the appropriate use of treatment consent and case difficulty assessment [29], which were listed in the core competencies in this study.

While CBME shows promise, implementation faces real challenges: resistance to change, budget constraints, and difficulty assessing competencies like attitudes. To succeed, institutions should involve stakeholders in redesign efforts, pursue external funding, and employ diverse assessment methods—including observation and simulation—to evaluate competency development more comprehensively [35].

According to a systematic review, questionnaires are the most widely used tool for identifying required competencies in medical education worldwide [24]. This aligns with the approach taken in this study.

The Delphi technique is among the most widely used techniques for defining competencies in medical education, offering many advantages over other decision-making methods. It stimulates a sense of responsibility and supports acceptance of consensus outcomes, minimizing bias from dominant individuals. This enhances the reliability and validity of the required core competencies [15,25]. A modified Delphi method incorporating consultation meetings with endodontic consultants was preferred for its accuracy and efficiency in identifying initial competencies and selecting experts. These meetings were conducted individually to prevent the influence of more senior or dominant consultants on others' opinions. The survey's results validated the consultation group, as all 27 excluded competencies were rated within the "important" range, except 3, which were in the "unsure" range. Notably, none of the 4 competencies proposed by participants in the first round were approved in the second round.

The diverse backgrounds of the participants contributed to a well-rounded and unbiased consensus on core endodontics competencies tailored to the needs of a Syrian society. Despite this diversity, the low standard deviation across both questionnaires indicates a strong agreement and minimal variation in opinions. The small standard error demonstrates that the mean ratings of the experts' opinions closely matched the approved means (Table 2). Finally, Cronbach α results confirm the internal consistency and reliability of the survey in both rounds (Table 3).

Medical literature suggests a sample size of approximately 20 individuals for this type of study [36], though larger sample sizes tend to yield more accurate results [37]. In this study, the sample size was sufficiently large, with at least 38 participants, whereas most similar studies typically involve 15-30 participants [38].

While multi-round studies with large samples often struggle with declining response rates—potentially undermining consensus credibility—our study achieves more than 70% response rate (38 of 53 participants), which meets or exceeds the standard benchmark many researchers consider essential for robust consensus findings [39,40].

The use of an electronic questionnaire via Google Forms was chosen to save time, effort, and cost. Nair et al [41] suggested strategies to improve survey response rates, including sending reminders at specific time intervals, an approach that was successfully implemented in this study, resulting in significant increases in responses after each reminder. Two rounds were adopted because no new competencies were added in the second round, and less than 9% of competencies were excluded. Jünger et al [42] stated in his methodological study that most Delphi-based research is satisfied with 2 rounds, which aligns with the procedure followed in this study. The threshold for accepting a competency as essential (must know) was set at a mean score above 4.20. Jünger et al [42] explained that the researcher determines this cutoff based on the nature and objectives of the research. This approach is consistent with similar studies identifying core competencies within specific medical specialties [18,32,40].

Twenty competencies, representing 37% of those proposed by the researcher and consultants, were excluded after the first round—a notably large proportion. This is because the primary references used to develop these competencies, especially as reported by Baaij et al [28] from the ESE, do not differentiate between the basic competencies required in undergraduate versus postgraduate endodontics, as the vast majority of



XSL•FO
RenderX

European countries do not recognize endodontics as a separate, independent specialty, but instead, they consider it as part of a general dentist's responsibilities [43,44]. Furthermore, these competencies were intended as recommendations for undergraduate endodontics curricula rather than essential competencies that students should master before graduation.

The [28] framework established 3 competency levels: be competent, have knowledge, and be familiar with. This study mirrors most ”be competent“ items, with 2 notable exceptions: managing uncertainty and maintaining pulp vitality were rated as “important“ by our expert panel. This distinction reflects expert opinion that these advanced competencies fall outside the general dentistry scope and are more appropriately reserved for endodontic specialists.

It is worth noting that 85% of the excluded competencies were still rated within the “important” category. These can be interpreted as important competencies for the student, and ”it is good to know,“ but they are not essential to master them at the undergraduate level.

By the end of the second round, 7 competencies were further excluded, including all 4 additional competencies suggested by participants. The remaining 3 competencies were close to the inclusion threshold for core competencies (Table 2).

Despite extensive research on identifying competencies in endodontics, no prior studies have directly addressed this topic comprehensively [29]. However, several institutions involved in medical and dental education such as the American Association of Endodontics, Association of Canadian Faculties of Dentistry, Canadian Medical Education Directives for Specialists, Commission of Dental Accreditation of Canada, ESE, General Dental Council, and the Saudi Commission for Health Specialties have provided guidelines for designing undergraduate or residency-level endodontic curricula.

The ESE recommended essential undergraduate endodontic competencies [28]. While this study identifies competencies that align closely with ESE's recommendations, the division of competency domains and the method of determining the competencies differ significantly. The ESE did not use a questionnaire nor distinguish between undergraduate and postgraduate endodontics, reflecting the absence of endodontic specialization in most European countries.

The study has some limitations that should be considered. Involving additional stakeholders such as students, patients, and health care or education administrators, through interviews rather than questionnaires, could enrich the findings but would also broaden the study's scope and complicate data analysis. So it would be preferable to build upon the results of this study, which relied on the endodontists’ opinions, with further studies targeting the views of other stakeholders to determine their needs and perspectives based on these findings.

Additionally, assigning numerical values 1-5 to responses may not capture the nuances of expert input; for example, a professor with more than 30 years of experience carries the same weight as a recent graduate. Nevertheless, including a range of perspectives across generations is important to ensure balanced and unbiased results. The high degree of harmony in opinions among the experts’ panel indicates an implicit agreement on the core competencies that a graduate should possess as a graduation requirement.

## Conclusion

This study identifies the essential endodontics competencies for undergraduate education in Syria by using a modified Delphi method. Further studies involving a broader range of stakeholders should be considered to confirm the results. Additionally, developing an international framework for undergraduate endodontics competencies through collaboration with researchers worldwide would be beneficial. The final list of the identified competencies provides a framework for developing a formal competency-based endodontics curriculum for the undergraduate stage, which can assist students in achieving the required standards and advancing the adoption of CBME in Syria and other similar settings.

A similar study must be conducted to identify the core competencies required for the postgraduate stage in endodontics in Syria. This would clearly differentiate between undergraduate and postgraduate endodontic requirements.

This study employs a modified Delphi method to identify essential endodontics competencies for undergraduate dental education in Syria. The resulting competency framework has been established as a foundation for developing a competency-based endodontics curriculum that supports students in meeting established standards and advancing CBME adoption across Syria and comparable contexts.

To strengthen these findings, broader stakeholder engagement is needed to validate the competency list. It is recommended that the research scope be expanded by including diverse stakeholders to confirm and refine results. Additionally, an international competency framework should be established through collaborative efforts with global researchers to ensure consistency and align best practices across regions.

Implementation of this framework should incorporate continuous assessment strategies, including clinical observation; multi-source feedback from patients, peers, and students; and learner self-assessment to monitor competency development and identify areas for improvement. These varied assessment methods enable timely intervention and support the transition toward CBME in Syria and similar settings.

Finally, a parallel study should be conducted to identify core competencies for postgraduate endodontics training in Syria. This research would create clear differentiation between undergraduate and graduate-level requirements, establishing a coherent progression pathway for dental professionals and ensuring comprehensive competency development across all training stages.



XSL•FO
RenderX

## Acknowledgments

All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article. The authors would also like to thank the Syrian Virtual University for its support, as well as all participants who provided comments and suggestions that led to this study. No generative artificial intelligence tools or assisted technologies were used to generate this manuscript.

## Data Availability

The datasets generated and analyzed during this study are not publicly available due to participant confidentiality and institutional restrictions. However, deidentified data may be made available from the corresponding author upon reasonable request and with appropriate ethical clearance.

## Funding

The authors declared that no financial support was received for this work.

## Authors' Contributions

MS is the principal researcher, collected participants' data, and analyzed the results. Both IJ and MD supervised the study; participated in the design and implementation of the study; and contributed to the study design, execution, analysis, and interpretation. The authors reviewed and agreed on all versions of the article before submission, agreed on the journal to which the article was submitted, and approved the final manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
First-round questionnaire: identification of core competencies required in endodontics for undergraduate students in Syrian dental schools, using a modified Delphi technique.
[DOCX File , 27 KB - ijmr_v15i1e83799_app1.docx ]

Multimedia Appendix 2
Second-round questionnaire: identification of core competencies required in endodontics for undergraduate students in Syrian dental schools, using a modified Delphi technique.
[DOCX File , 26 KB - ijmr_v15i1e83799_app2.docx ]

## References

1. Alrahabi M, Ahmad M. Knowledge regarding technical aspects of non-surgical root canal treatment in Al-Madinah Al-Munawarah private dental centers. Saudi Endod J 2015;5(3):155-160. [doi: 10.4103/1658-5984.163625]
2. Dietz GC S, Dietz GC J. The endodontist and the general dentist. Dental Clinics of North America 1992 Apr;36(2):459-471. [doi: 10.1016/s0011-8532(22)02507-1] [Medline: 1533378]
3. Al-Obaida M, Alwehaiby K, Al-Hindi O, Merdad K, Al-Madi E. Radiographic evaluation of the technical quality of root canal filling in Riyadh government and private hospitals. Saudi Endod J 2020 Apr;10(3):194-198. [doi: 10.4103/sej.sej_156_19] [Medline: 1533378]
4. Alrahabi M, Younes H. A cross-sectional study of the quality of root canal treatment in Al-Madinah Al-Munawwarah. Saudi Endod J 2016 Apr;6(1):31-35. [doi: 10.4103/1658-5984.172005]
5. Di Filippo G, Sidhu SK, Chong BS. Apical periodontitis and the technical quality of root canal treatment in an adult sub-population in London. Br Dent J 2014 May;216(10):E22. [doi: 10.1038/sj.bdj.2014.404] [Medline: 24853011]
6. Chueh L, Chen S, Lee C, Hsu Y, Pai S, Kuo M, et al. Technical quality of root canal treatment in Taiwan. Int Endod J 2003 Jun;36(6):416-422. [doi: 10.1046/j.1365-2591.2003.00672.x] [Medline: 12801289]
7. Boucher Y, Matossian L, Rilliard F, Machtou P. Radiographic evaluation of the prevalence and technical quality of root canal treatment in a French subpopulation. Int Endod J 2002 Mar;35(3):229-238. [doi: 10.1046/j.1365-2591.2002.00469.x] [Medline: 11985674]
8. Koch M, Wolf E, Tegelberg A, Petersson K. Effect of education intervention on the quality and long-term outcomes of root canal treatment in general practice. Int Endod J 2015 Jul;48(7):680-689. [doi: 10.1111/iej.12367] [Medline: 25112721]
9. Qualtrough AJE. Undergraduate endodontic education: what are the challenges? Br Dent J 2014 Mar;216(6):361-364. [doi: 10.1038/sj.bdj.2014.227] [Medline: 24651348]
10. Commission on Dental Accreditation Standards for Dental Education Programs. Commission on Dental Accreditation. 2018 Jan. URL: https://coda.ada.org/standards [accessed 2021-07-08]



XSL•FO
RenderX

11. Levin L, Halperin-Sternfeld M. Tooth preservation or implant placement: a systematic review of long-term tooth and implant survival rates. J Am Dent Assoc 2013;144(10):1119-1133. [doi: 10.14219/jada.archive.2013.0030] [Medline: 24080928]
12. Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, et al. Competency-based medical education: theory to practice. Med Teach 2010;32(8):638-645. [doi: 10.3109/0142159X.2010.501190] [Medline: 20662574]
13. ten Cate O, Scheele F. Viewpoint: competency-based postgraduate training: can we bridge the gap between theory and clinical practice? Academic Medicine 2007;82(6):542-547 [FREE Full text] [doi: 10.1097/acm.0b013e31805559c7] [Medline: 17525536]
14. Frank JR, Mungroo R, Ahmad Y, Wang M, De Rossi S, Horsley T. Toward a definition of competency-based education in medicine: a systematic review of published definitions. Med Teach 2010;32(8):631-637. [doi: 10.3109/0142159X.2010.500898] [Medline: 20662573]
15. Shah S, McCann M, Yu C. Developing a national competency-based diabetes curriculum in undergraduate medical education: a Delphi study. Can J Diabetes 2020 Feb;44(1):30-36. [doi: 10.1016/j.jcjd.2019.04.019] [Medline: 31399366]
16. Veale P, Busche K, Touchie C, Coderre S, McLaughlin K. Choosing our own pathway to competency-based undergraduate medical education. Acad Med 2019 Jan;94(1):25-30. [doi: 10.1097/ACM.0000000000002410] [Medline: 30113362]
17. Souliman M, Dashash M. Reasons for failure to change Syrian medical curricula: a qualitative descriptive study. Azerbaijan Medical Journal 2023;63:8407-8414 [FREE Full text]
18. Zoukar I, Dashash M. Using a modified Delphi method for identifying competencies in a Syrian undergraduate neonatology curriculum. Matern Child Health J 2023 Nov;27(11):1921-1929. [doi: 10.1007/s10995-023-03719-z] [Medline: 37289293]
19. Soundariya K, Nishanthi A, Mahendran R, Vimal M. Evaluation of competency-based medical education (CBME) curriculum implementation for phase II medical undergraduates: a qualitative study. J Adv Med Educ Prof 2025 Jan;13(1):36-48. [doi: 10.30476/jamp.2024.104312.2052] [Medline: 39906076]
20. Leung WC. Competency based medical training: review. BMJ 2002 Sep 28;325(7366):693-696 [FREE Full text] [doi: 10.1136/bmj.325.7366.693] [Medline: 12351364]
21. Danilovich N, Kitto S, Price DW, Campbell C, Hodgson A, Hendry P. Implementing competency-based medical education in family medicine: a narrative review of current trends in assessment. Fam Med 2021 Jan;53(1):9-22 [FREE Full text] [doi: 10.22454/FamMed.2021.453158] [Medline: 33471918]
22. Caccia N, Nakajima A, Kent N. Competency-based medical education: the wave of the future. J Obstet Gynaecol Can 2015 Apr;37(4):349-353. [doi: 10.1016/S1701-2163(15)30286-3] [Medline: 26001689]
23. Hawkins RE, Welcher CM, Holmboe ES, Kirk LM, Norcini JJ, Simons KB, et al. Implementation of competency-based medical education: are we addressing the concerns and challenges? Med Educ 2015 Nov;49(11):1086-1102. [doi: 10.1111/medu.12831] [Medline: 26494062]
24. Alharbi N. Evaluating competency-based medical education: a systematized review of current practices. BMC Med Educ 2024 Jun 03;24(1):612 [FREE Full text] [doi: 10.1186/s12909-024-05609-6] [Medline: 38831271]
25. Dashash M, Almasri B, Takaleh E, Halawah A, Sahyouni A. Educational perspective for the identification of essential competencies required for approaching patients with COVID-19. East Mediterr Health J 2020 Sep 24;26(9):1011-1017 [FREE Full text] [doi: 10.26719/emhj.20.111] [Medline: 33047791]
26. Geng Y, Zhao L, Wang Y, Jiang Y, Meng K, Zheng D. Competency model for dentists in China: results of a Delphi study. PLoS One 2018;13(3):e0194411 [FREE Full text] [doi: 10.1371/journal.pone.0194411] [Medline: 29566048]
27. Benzian H, Greenspan J, Barrow J, Hutter J, Loomer P, Stauf N, et al. A competency matrix for global oral health. Journal of Dental Education 2015 Apr;79(4):353-361 [FREE Full text] [doi: 10.1002/j.0022-0337.2015.79.4.tb05891.x] [Medline: 25838005]
28. Baaij A, Kruse C, Whitworth J, Jarad F. European Society of Endodontology undergraduate curriculum guidelines for endodontology. Int Endod J 2024 Aug;57(8):982-995. [doi: 10.1111/iej.14064] [Medline: 38551606]
29. Algahtani FN, Barakat RM, Almohareb RA, Alqarni L, Alqabbani A, Almadi E. The objectives and instructional design of undergraduate endodontic program: multicenter cross-sectional study in Saudi Arabia. BMC Med Educ 2022 Jun 23;22(1):486 [FREE Full text] [doi: 10.1186/s12909-022-03548-8] [Medline: 35733185]
30. ACFD educational framework for the development of competency in dental programs. Association of Canadian Faculties of Dentistry. 2016. URL: https://acfd.ca/wp-content/uploads/ACFD-Educational-Framework.pdf [accessed 2025-03-02]
31. Guan L, Gao P, Liu S, Liu Y, Li X, Liu F, et al. Development of a global health bachelor curriculum in China: a Delphi study. BMJ Open 2019 Jan 03;9(1):e023893 [FREE Full text] [doi: 10.1136/bmjopen-2018-023893] [Medline: 30610023]
32. Zoukar I, Zoukar S, Muad M, Dashash M. Establishing core competencies for neonatology fellowship training in Syria: a modified Delphi study approach. BMC Med Educ 2025 Jul 28;25(1):1123 [FREE Full text] [doi: 10.1186/s12909-025-07696-5] [Medline: 40722172]
33. Dutner JM, Sidow SJ, Cervero RM, Soh M. Endodontic competence of the new general dentist as conceptualized by predoctoral educators: a qualitative exploration. J Dent Oral Epidemiol 2024 Mar 27;4(2):JDOE2400109 [FREE Full text] [doi: 10.54289/jdoe2400109]




XSL•FO
RenderX

34. Algahtani FN, Barakat RM, Alqarni LM, Alqabbani AF, Alkadi MF, Almohareb RA. Undergraduate endodontic training and its relation to contemporary practice: multicenter cross-sectional study in Saudi Arabia. Int J Clin Pract 2023 Jan;2023:7484570 [FREE Full text] [doi: 10.1155/2023/7484570] [Medline: 36704246]
35. Ramanathan R, Shanmugam J, Gopalakrishnan S, Palanisamy K, Narayanan S. Challenges in the implementation of competency-based medical curriculum: perspectives of prospective academicians. Cureus 2022 Dec;14(12):e32838 [FREE Full text] [doi: 10.7759/cureus.32838] [Medline: 36694520]
36. Dunn WR, Hamilton DD, Harden RM. Techniques of identifying competencies needed of doctors. Med Teach 1985;7(1):15-25. [doi: 10.3109/01421598509036787] [Medline: 3913836]
37. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000 Oct;32(4):1008-1015. [doi: 10.1046/j.1365-2648.2000.01567.x] [Medline: 11095242]
38. Sahyouni A, Zoukar I, Dashash M. Evaluating the effectiveness of an online course on pediatric malnutrition for Syrian health professionals: qualitative delphi study. JMIR Med Educ 2024 Oct 28;10:e53151 [FREE Full text] [doi: 10.2196/53151] [Medline: 39651700]
39. Keeney S, McKenna H, Hasson F. The Delphi Technique in Nursing and Health Research. West Sussex, United Kingdom: John Wiley & Sons Ltd; 2011.
40. Giannarou L, Zervas E. Using Delphi technique to build consensus in practice. IJBSAM 2014;9(2):65-82. [doi: 10.69864/ijbsam.9-2.106]
41. Nair CS, Wayland C, Soediro S. Evaluating the student experience: a leap into the future. 2005 Nov 28 Presented at: Australasian Evaluations Forum: University Learning and Teaching; November 28-29; Sydney, Australia URL: https://tinyurl.com/4exwjp65
42. Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: recommendations based on a methodological systematic review. Palliat Med 2017 Sep;31(8):684-706. [doi: 10.1177/0269216317690685] [Medline: 28190381]
43. García-Espona I, García-Espona E, Alarcón JA, Fernández-Serrano J. European inequalities and similarities in officially recognized dental specialties. BMC Oral Health 2023 May 11;23(1):280 [FREE Full text] [doi: 10.1186/s12903-023-02987-z] [Medline: 37170254]
44. Al Raisi H, Dummer PMH, Vianna ME. How is endodontics taught? A survey to evaluate undergraduate endodontic teaching in dental schools within the United Kingdom. Int Endod J 2019 Jul;52(7):1077-1085. [doi: 10.1111/iej.13089] [Medline: 30706491]


## Abbreviations

**CBME:** competency-based medical education
**ESE:** European Society of Endodontology


*Edited by M Balcarras; submitted 09.Sep.2025; peer-reviewed by L Alsofi, D Winterton; comments to author 15.Apr.2026; revised version received 26.May.2026; accepted 26.May.2026; published 09.Jun.2026.*

*Please cite as:*
*Salameh M, Dashash M, Jamous I*
*Identification of the Core Competencies Required in Endodontics for Undergraduate Students in Syrian Dental Schools by Using a Modified Delphi Technique: Prospective Exploratory Survey Study*
*Interact J Med Res 2026;15:e83799*
*URL: https://www.i-jmr.org/2026/1/e83799*
*doi:10.2196/83799*
*PMID:*


©Muhammad Salameh, Mayssoon Dashash, Issam Jamous. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 09.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# COVID-19 Rebound in Nirmatrelvir Plus Ritonavir Treatment and Control Groups: Prospective Cohort Study

Jacqueline K Kueper[1], PhD; Kalyani Kottilil[1], BSc; Giorgio Quer[1], PhD; Danielle C Chiang[1], MPH; Emily G Spencer[1], PhD; Jyothi Purushotham[1,2], PhD; Edward Ramos[1,3], PhD; Leila Roumani[4], DMD, MPH; Kristian G Andersen[1,2], PhD; Eric J Topol[1], MD; Jay A Pandit[1*], MD; Michael J Mina[4,5*], MD, PhD

[1]Scripps Research Translational Institute, Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA, United States

[2]Department of Immunology and Microbiology, Scripps Research Institute, La Jolla, CA, United States

[3]Care Evolution, Ann Arbor, MI, United States

[4]eMed, Miami, FL, United States

[5]Immune Observatory, Boston, MA, United States

[*]these authors contributed equally

**Corresponding Author:**
Jacqueline K Kueper, PhD
Scripps Research Translational Institute, Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA, United States

## Abstract

**Background:** Observation of COVID-19 rebound after nirmatrelvir plus ritonavir (NPR) has driven important questions surrounding one of the only direct-acting antiviral treatments for COVID-19.

**Objective:** The objective of this study was to examine the epidemiology of COVID-19 rebound among COVID-19–positive outpatients in the United States who independently decided whether or not to take NPR.

**Methods:** This prospective, decentralized observational cohort study was conducted from August 2022 through December 2023 and included frequent proctored COVID-19 rapid antigen tests and self-report symptom surveys for 15 days. The primary outcome was the incidence of viral and symptom rebound. Secondary outcomes included time to initial viral and symptom clearance, rebound probability among patients who cleared by day 15, and symptom frequency.

**Results:** Of 917 consenting participants, 669 (73%) were eligible for inclusion in the analysis (n=443, 66% in the NPR group; n=226, 34% in the control group). The mean age was 46.1 (SD 12.9) years, 62.6% (n=419) of participants were female, and 49.2% (n=329) had at least one preexisting condition. Overall, 15-day cumulative incidence was higher in the NPR group than the control group for both viral (70/443, 15.8% vs 12/226, 5.3%) and symptom (73/443, 16.5% vs 19/226, 8.4%) rebound. Time to initial viral and symptom clearance was similar between groups, and among those who experienced clearance by day 15, the probability of viral rebound (NPR: 19.1%, 95% CI 15.1%-24.0% vs control: 7%, 95% CI 4.0%-12.6%; *P*<.001) and symptom rebound (NPR: 47.7%, 95% CI 36.1%-60.8% vs control: 16.9%, 95% CI 10.9%-25.7%; *P*<.001) was higher in the NPR group than the control group.

**Conclusions:** This study demonstrates that while COVID-19 rebound occurs in both NPR-treated and untreated outpatients, the incidence is higher in the NPR group.

***(Interact J Med Res 2026;15:e80263)*** doi:10.2196/80263

**KEYWORDS**

COVID-19; COVID-19 rebound; nirmatrelvir plus ritonavir; symptom rebound; viral rebound

## Introduction

In the United States, more than 1 million deaths and approximately 7 million hospitalizations reflect the toll of SARS-CoV-2 [1]. After initial emergency use authorization in December 2021 [2], nirmatrelvir plus ritonavir (NPR; Paxlovid) became the first fully approved oral antiviral treatment by the US Food and Drug Administration in May 2023 [3]. US COVID-19 treatment guidelines align on a strong recommendation for a 5-day course of NPR in outpatients at high risk of progression to severe COVID-19 [4-6].

The guidance in favor of NPR is largely based on studies showing significant protective effects against hospitalization and death for high-risk outpatients, with little evidence of serious side effects, across changing COVID-19 variants and baseline population risk levels [4-7]. Despite this, it has been estimated that less than a quarter of eligible patients use NPR, potentially missing opportunities to further reduce serious illness and death



XSL•FO
RenderX

rates [8-10]. Hesitation in NPR prescription may be associated with anecdotal reports and poor knowledge about COVID-19 rebound [8,11].

COVID-19 rebound is the resurgence in viral load and/or COVID-19 symptoms after initial recovery. This phenomenon first gained attention as a secondary finding in early studies testing NPR effectiveness against severe disease outcomes, although it also occurs in untreated individuals [12,13]. In direct response to the need for further investigation, we opened a prospective observational cohort in 2022 (COVID-19 Rebound Study) to evaluate the epidemiology of viral and symptom rebound differences in NPR treatment and control individuals based on high-frequency testing and symptom surveys [14]. Preliminary results found higher than previously reported COVID-19 viral and symptom rebound in both NPR and control groups [14]. While the amount of research on rebound is increasing, evidence supporting a potential association with NPR treatment is still relatively weak, relying on retrospective chart reviews or secondary analyses [15].

This paper describes the final results of the COVID-19 Rebound Study, expanding upon our previously published preliminary analysis [14], with a larger sample size and extended enrollment period to evaluate the epidemiology of COVID-19 rebound in participants who were clinically eligible for NPR and who did or did not decide to take NPR.

## Methods

### Study Design and Recruitment

This decentralized, prospective observational study was completed via collaboration between the Scripps Research Translational Institute (SRTI) and eMed, which implemented the National Institutes of Health Home Test to Treat Program, a 24/7 virtual health platform that provides proctor-guided at-home testing with public health reporting through a digital Clinical Laboratory Improvement Amendments (CLIA)–waived laboratory. Eligibility criteria included a positive rapid antigen test for COVID-19 within 48 hours prior to study enrollment, residing within the United States, speaking English, and aged at least 18 years. eMed users who tested positive were offered NPR at the discretion of an eMed provider. NPR was the only COVID-19 oral antiviral therapy offered through the eMed platform. There were two study recruitment cohorts: (1) from August 2022 through December 2023, eligible eMed users that tested positive for COVID-19 through an eMed telehealth appointment were invited via follow-up email, text, and/or phone call; and (2) from August 2023 through December 2023, social media campaigns and SRTI blog posts invited people who self-tested positive to consider participation. Participation required a virtual meeting with an eMed provider to confirm a positive COVID-19 rapid antigen test and for further eligibility screening.

### Study Procedures

Interested individuals were sent to a landing page with more information on the study, an eligibility screener survey, and e-consent forms. Upon providing consent, participants reported baseline information (self-reported demographic characteristics, preexisting conditions, and vaccination history) and completed a symptom survey prior to requesting a study kit. Study kits were shipped overnight and included detailed instructions for the 15-day acute study phase and eMed live-proctor–enabled COVID-19 test-to-treat rapid antigen test kits (containing Abbott BinaxNOW Home tests). Participants who enrolled after August 2023 (cohort 2) were additionally sent 2 at-home blood collection devices with biohazard shipping materials. Cohort 2 was instructed to return blood samples and testing swabs to SRTI for serological analyses and viral genome sequencing. Results from blood sampling and sequencing are forthcoming in a future report.

Participants were asked to complete a survey (capturing symptoms and NPR use) and a proctored eMed COVID-19 rapid antigen test on days 2, 5, 7, 9, 11, 13, and 15 from enrollment (acute study phase). To complete a rapid antigen COVID-19 test, participants were instructed to scan a QR code on their test box to initiate a telehealth proctored visit where they were guided through the testing procedure and test results validated. If a participant tested positive during the visit and had not already received care either in person or through a telemedicine visit, they were offered an optional eMed telemedicine visit that could provide a prescription for a 5-day course of NPR (provided under the brand name Paxlovid). Postacute COVID-19 symptom surveys were sent to all participants at 1-, 3-, and 6-months after enrollment, with results forthcoming. Participants received up to 3 notification reminders per task using their preferred mode of communication (email, text, and phone) unless they chose to withdraw early. Research Electronic Data Capture (REDCap) tools were used for survey and notification administration [16,17]. Participants were offered an incentive of electronic gift cards totaling up to US $200, which were issued over the study period according to completion of required tasks.

### Outcomes

The primary outcomes were cumulative incidence of viral and symptom rebound in NPR and control groups within 15 days of enrollment. *Viral rebound* was defined as a positive rapid antigen test after previously testing negative. *Symptom rebound* was defined as reporting at least one symptom on a survey after reporting zero symptoms on a previous survey. Secondary outcomes included time from enrollment to initial negative rapid antigen test and to symptom resolution (viral and symptom clearance, respectively), and rebound probability among patients with clearance by day 15. Exploratory outcomes included comparing characteristics of rebound and nonrebound groups and symptom frequency during the acute study phase. Reported symptoms were aggregated into binary categories of systemic, respiratory, gastrointestinal, neurological, and other symptoms.

### Statistical Analyses

#### *Treatment Groups*

Participants who reported their independent decision to take at least one NPR dose were classified as the NPR group, and remaining participants were classified as the control group. Participants were excluded if they did not complete the baseline survey or at least 2 follow-up surveys, or if they had missing



XSL•FO
RenderX

data precluding ascertainment of rebound. Participants reporting no symptoms at baseline (n=5) were excluded from symptom rebound analyses.

### *Main Analyses*

Cumulative incidence was calculated as the number of rebound events across the 15-day acute study phase divided by the number of eligible participants. Kaplan-Meier survival curves were generated and stratified by NPR and control groups to compare timing of rebound since study enrollment, of viral and symptom clearance (time from positive test or symptoms to first negative test or no symptoms), and of rebound probability among patients who cleared by day 15 (time from negative test or no symptoms to subsequent positive test or symptoms). In each time-to-event analysis, participants were censored at the time of missing more than 2 consecutive surveys or rapid antigen tests or after day 15 if they did not have an event. The 95% CIs were calculated using the Greenwood exponential formula.

### *Secondary Analyses*

Due to potential differences between participants recruited via the eMed platform and social media, we conducted sensitivity analyses comparing baseline characteristics and outcomes between recruitment cohorts. We additionally conducted 5 secondary analyses. First, we compared baseline participant characteristics, rebound incidence, and symptom frequency between the 2 cohorts (ie, recruited before or after August 2023). Second, we assessed sensitivity of cumulative incidence estimates to different censoring (restricted to those with no missing data or those with at least 3 surveys before missing any data), time (11- vs 15-day estimates), and amount of NPR use (none, 1 - 4 days, or 5 days) criteria. Participant characteristics and frequency of symptom surveys were additionally compared between groups defined by the amount of NPR use reported. Third, we adjusted rebound incidence comparisons for baseline characteristics using Cox proportional hazards regression models. Separate models were run separately for viral and symptom rebound and included covariates for age, number of vaccines, sex, race, and preexisting conditions. As compared with the main participant characteristic table groupings, additional data cleaning was performed to collapse small cell counts, including grouping of cardiovascular conditions (heart disease, heart failure, and high blood pressure), grouping of lung conditions (asthma, chronic bronchitis, chronic obstructive pulmonary disease, emphysema, and other lung conditions), removing the single intersex participant, and combining race categories with fewer than 3 participants into the other category. Fourth, we examined congruence between symptom and test positivity by counting the number of person-days (ie, unique participant surveys) where symptoms and tests were both positive, negative, or opposite of each other. This was done for all acute phase days and for the subset of person-days where a participant was testing positive in an episode of viral test positivity. Fifth, we compared baseline characteristics of participants who were and were not eligible for inclusion in analyses.

Analyses were done with Python (version 3.8; Python Software Foundation). Between-group statistical comparisons of baseline characteristics and symptoms were performed using the TableOne package with 2-sided chi-square tests for categorical variables and *t* tests for continuous variables, alpha level .05 [18]. The Benjamini-Hochberg procedure to control for false discovery rate was applied across multiple tests. Time-to-event analyses were performed using the lifelines package [19]. We followed STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) reporting guidelines [20] (Multimedia Appendix 1).

### Ethical Considerations

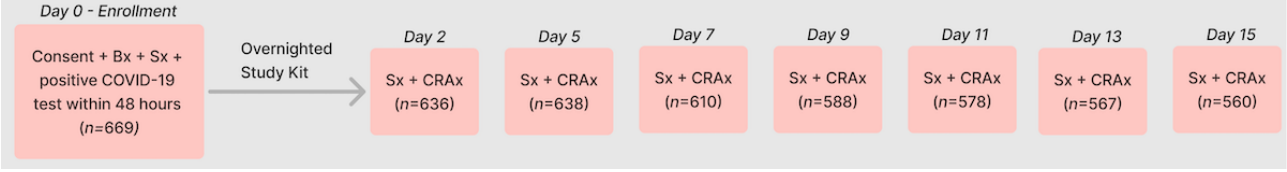
The study was approved by the Scripps Institutional Review Board (22 - 7978). All participants provided electronic informed consent. SRTI performed safety oversight.

## Results

### Study Population

Of 917 consented participants, 669 (73%) were eligible for inclusion in analyses (Figure 1); of these, 443 (66.2%) reported taking at least one dose of NPR (NPR group) and 226 (33.8%) did not (control group). Within the NPR group (n=443), 197 (44.5%) reported taking all 5 days of their NPR prescription. The average age was 46.1 (SD 12.9) years, and 62.6% (419/669) of participants were female. The most common preexisting condition was high blood pressure (132/669, 19.7%), and half of participants (340/669, 50.8%) did not report any preexisting conditions. While sex was similar between groups, there were differences in age (56/67, 83.6% of those older than 65 years opting to take NPR compared with 189/258, 73.3% of those aged 45-65 years and 198/344, 57.6% of those aged 18-44 years), race (351/491, 71.5% of White participants opting to take NPR compared with 85/168, 50.3% of non-White participants), and the presence of preexisting conditions (Table 1). The proportion of participants in the NPR group was higher in the first cohort (181/251, 72.1%) than the second cohort (262/418, 62.7%; Table S1 in Multimedia Appendix 1).

**Figure 1.** Study flow. Study task completion for the analysis sample across days since enrollment. Bx: baseline survey; CRAx: study provided COVID-19 rapid antigen test; n: number of participants meeting analysis eligibility criteria that completed the survey; Sx: symptom survey.






XSL•FO
RenderX

**Table .** Participant characteristics by treatment group[a].

| | Overall (n=669) | Control (n=226) | NPR[b] (n=443) | *P* value |
|---|---|---|---|---|
| Age (years), n (%) | | | | .005 |
| 18 - 44 | 344 (51.4) | 146 (64.6) | 198 (44.7) | |
| 45 - 64 | 258 (38.6) | 69 (30.5) | 189 (42.7) | |
| ≥65 | 67 (10) | 11 (4.9) | 56 (12.6) | |
| Sex, n (%) | | | | .87 |
| Female | 419 (62.6) | 143 (63.3) | 276 (62.3) | |
| Male | 249 (37.2) | 83 (36.7) | 166 (37.5) | |
| Intersex | 1 (0.1) | 0 (0) | 1 (0.2) | |
| Race, n (%) | | | | .005 |
| American Indian or Alaska Native | 2 (0.3) | 1 (0.4) | 1 (0.2) | |
| Asian | 42 (6.3) | 19 (8.4) | 23 (5.2) | |
| Black, African American, or African | 41 (6.1) | 20 (8.8) | 21 (4.7) | |
| Hispanic, Latino, or Spanish | 47 (7) | 27 (11.9) | 20 (4.5) | |
| White | 491 (73.4) | 140 (61.9) | 351 (79.2) | |
| Multiple | 33 (4.9) | 15 (6.6) | 18 (4.1) | |
| Other | 3 (0.4) | 1 (0.4) | 2 (0.4) | |
| Declined to answer | 10 (1.5) | 3 (1.3) | 7 (1.6) | |
| At least one COVID-19 vaccine, n (%) | 637 (95.2) | 208 (92.0) | 429 (96.8) | .04 |
| Preexisting conditions, n (%) | | | | |
| Asthma | 117 (17.5) | 34 (15) | 83 (18.7) | .52 |
| Autoimmune condition | 63 (9.4) | 17 (7.5) | 46 (10.4) | .52 |
| Cancer | 35 (5.2) | 5 (2.2) | 30 (6.8) | .06 |
| Chronic bronchitis | 10 (1.5) | 3 (1.3) | 7 (1.6) | >.99 |
| Chronic obstructive pulmonary disease | 3 (0.4) | 3 (1.3) | 0 (0) | .10 |
| Diabetes | 44 (6.6) | 10 (4.4) | 34 (7.7) | .34 |
| Emphysema | 1 (0.1) | 1 (0.4) | 0 (0) | .54 |
| Heart disease | 21 (3.1) | 5 (2.2) | 16 (3.6) | .56 |
| Heart failure | 3 (0.4) | 1 (0.4) | 2 (0.5) | >.99 |
| High blood pressure | 132 (19.7) | 40 (17.7) | 92 (20.8) | .56 |
| Other lung condition | 7 (1) | 1 (0.4) | 6 (1.4) | .56 |
| None | 340 (50.8) | 137 (60.6) | 203 (45.8) | .005 |
| 15-day COVID-19 rebound incidence, n (%) | | | | |
| Viral test rebound | 82 (12.3) | 12 (5.3) | 70 (15.8) | .002 |
| Symptom rebound | 92 (13.8) | 19 (8.4) | 73 (16.5) | .006 |
| Acute phase symptom surveys, mean (SD) | | | | |
| Systemic symptoms | 3.5 (2.4) | 3.2 (2.2) | 3.7 (2.4) | .04 |
| Respiratory symptoms | 4.7 (2.4) | 4.5 (2.4) | 4.8 (2.4) | .13 |
| Gastrointestinal symptoms | 1.0 (1.4) | 0.9 (1.3) | 1.1 (1.5) | .06 |
| Neurological symptoms | 0.4 (1.0) | 0.3 (0.7) | 0.5 (1.2) | .04 |
| Other symptoms | 0.4 (0.9) | 0.3 (0.8) | 0.4 (1.0) | .06 |

[a]The acute phase symptom category measures are cumulative and represent the number of survey days across the 15-day acute study phase during which participants reported experiencing each symptom type (maximum 8 surveys).



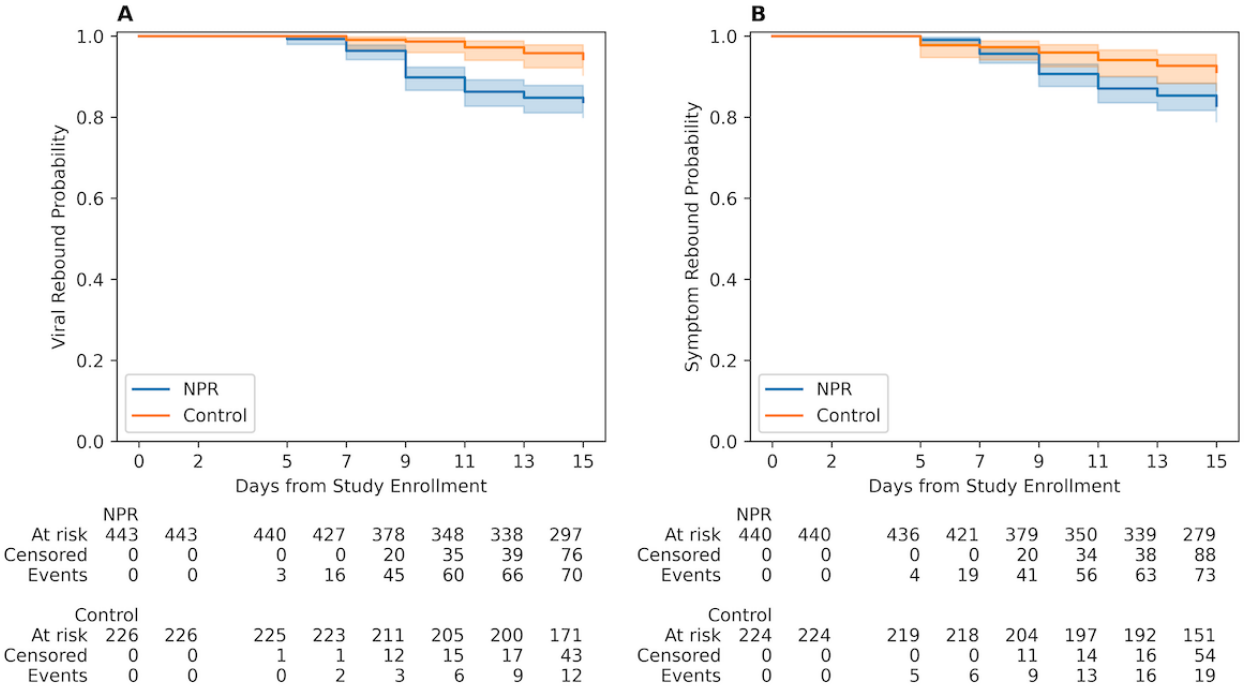
XSL•FO
RenderX

[b]NPR: nirmatrelvir plus ritonavir.

## Primary Outcome: Rebound Incidence

The 15-day cumulative viral rebound incidence was 12.3% (82/669) overall (NPR: 70/443, 15.8% and control: 12/226, 5.3%). The 15-day cumulative symptom rebound incidence was 13.8% (92/664) overall (NPR: 73/443, 16.5% and control: 19/226, 8.4%). Kaplan-Meier curves showed significantly increased probability of viral test rebound by day 15 of 16.2% (95% CI 13.1% - 20.1%) for the NPR group and 5.6% (95% CI 3.2% - 9.6%) for the control group (Figure 2; log-rank $P$<.001). Probability of symptom rebound by day 15 was 17.2% (95% CI 13.9% - 21.1%) and 8.8% (95% CI 5.7% - 13.4%) for NPR and control groups, respectively (Figure 2; log-rank $P$=.004). Secondary analyses examining incidence across different eligibility and timing criteria (Table S2 in Multimedia Appendix 1), whether a partial (1 - 4 days) or full (5 days) prescription of NPR was taken (Table S3 in Multimedia Appendix 1), and adjusted for baseline demographic and clinical characteristics using Cox proportional hazards regression (Tables S4 and S5 in Multimedia Appendix 1) did not change conclusions.

**Figure 2.** Kaplan-Meier curves between nirmatrelvir plus ritonavir (NPR) and control groups for overall rebound risk from baseline. Curves present the probability of events for our primary outcome measure of overall (A) viral and (B) symptom rebound risk since study enrollment.




## Secondary Outcomes

### *Initial Viral and Symptom Clearance*

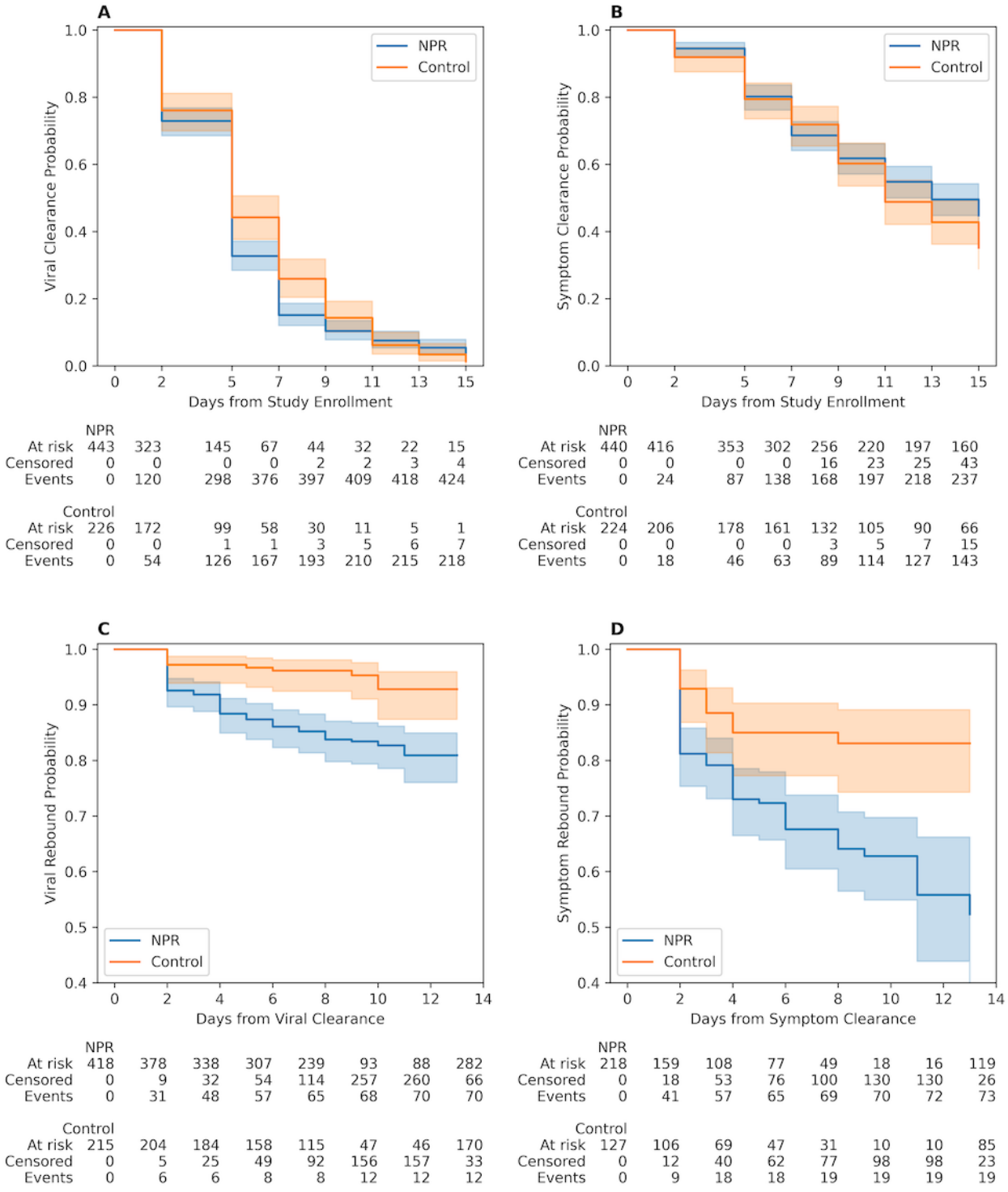
Time from enrollment to initial viral clearance was similar in the NPR and control groups (Figure 3; log-rank $P$=.11). Among the 669 participants, 642 (96%) who experienced viral clearance, the mean time to clearance was 5.4 (SD 2.9) days (NPR: mean 5.2, SD 2.8 days and control: mean 5.9, SD 3.1 days; Figure S1 in Multimedia Appendix 1). A total of 16 (2.4%) participants (NPR: 15/16, 93.7% and control: 1/16, 6.3%) never tested negative throughout the 15-day acute study phase. Another 11 (1.6%) participants (NPR: 4/11, 36.4% and control: 7/11, 63.6%) had unknown clearance timing due to too many sequentially missing tests.

Time to initial symptom clearance tended to take longer than viral clearance but was also similar between the NPR and control groups (Figure 3; log-rank $P$=.06). Of the 664 participants who reported at least one symptom at baseline, 380 (57.2%) reported symptom clearance in a mean of 8.1 (SD 3.8) days (NPR: mean 7.9, SD 3.7 days and control: mean 8.5, SD 4.0 days; Figure S1 in Multimedia Appendix 1). There were 226 (34%) participants (NPR: 160/226, 70.8% and control: 66/226, 29.2%) who did not report symptom clearance by day 15. An additional 58 (8.7%) participants (NPR: 43/58, 74.1% and control: 15/58, 25.9%) had too many sequentially missing symptom surveys to determine clearance timing.



XSL•FO
RenderX

**Figure 3.** Kaplan-Meier curves between nirmatrelvir plus ritonavir (NPR) and control groups for clearance and clearance-dependent rebound risk. Curves present the probability of events for our secondary outcomes of initial clearance of (A) virus test positivity or (B) resolution of symptoms and (C and D) risk of rebound following initial clearance or resolution.




| A | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NPR | | | | | | | | |
| At risk | 443 | 323 | 145 | 67 | 44 | 32 | 22 | 15 |
| Censored | 0 | 0 | 0 | 0 | 2 | 2 | 3 | 4 |
| Events | 0 | 120 | 298 | 376 | 397 | 409 | 418 | 424 |
| Control | | | | | | | | |
| At risk | 226 | 172 | 99 | 58 | 30 | 11 | 5 | 1 |
| Censored | 0 | 0 | 1 | 1 | 3 | 5 | 6 | 7 |
| Events | 0 | 54 | 126 | 167 | 193 | 210 | 215 | 218 |

| B | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NPR | | | | | | | | |
| At risk | 440 | 416 | 353 | 302 | 256 | 220 | 197 | 160 |
| Censored | 0 | 0 | 0 | 0 | 16 | 23 | 25 | 43 |
| Events | 0 | 24 | 87 | 138 | 168 | 197 | 218 | 237 |
| Control | | | | | | | | |
| At risk | 224 | 206 | 178 | 161 | 132 | 105 | 90 | 66 |
| Censored | 0 | 0 | 0 | 0 | 3 | 5 | 7 | 15 |
| Events | 0 | 18 | 46 | 63 | 89 | 114 | 127 | 143 |

| C | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NPR | | | | | | | | |
| At risk | 418 | 378 | 338 | 307 | 239 | 93 | 88 | 282 |
| Censored | 0 | 9 | 32 | 54 | 114 | 257 | 260 | 66 |
| Events | 0 | 31 | 48 | 57 | 65 | 68 | 70 | 70 |
| Control | | | | | | | | |
| At risk | 215 | 204 | 184 | 158 | 115 | 47 | 46 | 170 |
| Censored | 0 | 5 | 25 | 49 | 92 | 156 | 157 | 33 |
| Events | 0 | 6 | 6 | 8 | 8 | 12 | 12 | 12 |

| D | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| NPR | | | | | | | | |
| At risk | 218 | 159 | 108 | 77 | 49 | 18 | 16 | 119 |
| Censored | 0 | 18 | 53 | 76 | 100 | 130 | 130 | 26 |
| Events | 0 | 41 | 57 | 65 | 69 | 70 | 72 | 73 |
| Control | | | | | | | | |
| At risk | 127 | 106 | 69 | 47 | 31 | 10 | 10 | 85 |
| Censored | 0 | 12 | 40 | 62 | 77 | 98 | 98 | 23 |
| Events | 0 | 9 | 18 | 18 | 19 | 19 | 19 | 19 |

### *Rebound Following Initial Clearance*

The primary outcome was rebound incidence among all participants under evaluation, regardless of whether they experienced initial viral or symptom clearance. Therefore, the previous evaluation provided an estimate of risk for a patient group that did not yet know if they would clear the virus or symptoms in time to become “eligible” to experience a rebound event within the 15-day acute phase. Thus, we also explored the critical question of rebound probability among patients who cleared by day 15.

Among the 633 participants who experienced initial viral clearance before day 15, the risk of viral rebound by day 15 was 19.1% (95% CI 15.1% - 24.0%) for the NPR group compared with 7.2% (95% CI 4.0% - 12.6%) for the control group (Figure



XSL•FO
RenderX

3; log-rank *P*<.001). The mean number of days from initial viral clearance to rebound positive test was 4.3 (SD 2.7) overall (NPR: mean 4.1, SD 2.5 days and control: 5.2, SD 3.5 days; Figure S2 in Multimedia Appendix 1). Almost one-third (139/418, 30%) of viral rebound cases in the NPR group lasted 2 days or less, while 17.1% (71/418), 11.4% (48/418), and 2.9% (12/418) lasted 2 to 4, 4 to 6, and 6 to 8 days, respectively (Figure S3 in Multimedia Appendix 1). A total of 38.6% (27/70) NPR and 3.6% (4/110) control group viral rebound cases did not reclear prior to the end of the acute phase follow-up or when they were censored for missing tests.

Among the 345 participants who experienced initial symptom clearance before day 15, the risk of symptom rebound by day 15 was 47.7% (95% CI 36.1% - 60.8%) for the NPR compared with 16.9% (95% CI 10.9% - 25.7%) for the control group (Figure 3; log-rank *P*<.001). The mean number of days from initial symptom resolution to rebound was 3.5 (SD 2.3) overall (NPR: mean 3.6, SD 2.5 days and control: mean 3.0, SD 1.4 days; Figure S2 in Multimedia Appendix 1). The duration of rebound symptoms is presented in Figure S3 in Multimedia Appendix 1, with 54.8% (40/73) of NPR and 47.4% (9/19) of control group symptom rebound cases not resolved by day 15 or when they were censored.

## Exploratory Analyses

### *Characteristics of the Rebound and Nonrebound Groups*

Table 2 compares characteristics of participants that did and did not experience each type of rebound. Of the 669 participants, there were 26 (3.9%) who experienced both viral and symptom rebound.



XSL•FO
RenderX

**Table .** Participant characteristics by viral and symptom test rebound[a].

| | No viral rebound (n=587) | Viral rebound (n=82) | *P* value | No symptom rebound (n=577) | Symptom rebound (n=92) | *P* value |
|---|---|---|---|---|---|---|
| Age (years), n (%) | | | .02 | | | .59 |
| 18 - 44 | 307 (52.3) | 37 (45.1) | | 298 (51.6) | 46 (50) | |
| 45 - 64 | 232 (39.5) | 26 (31.7) | | 225 (39) | 33 (35.9) | |
| ≥65 | 48 (8.2) | 19 (23.2) | | 54 (9.4) | 13 (14.1) | |
| Sex, n (%) | | | >.99 | | | .26 |
| Female | 366 (62.4) | 53 (64.6) | | 368 (63.8) | 51 (55.4) | |
| Male | 220 (37.5) | 29 (35.4) | | 209 (36.2) | 40 (43.5) | |
| Intersex | 1 (0.2) | 0 (0) | | 0 (0) | 1 (1.1) | |
| Race, n (%) | | | >.99 | | | .50 |
| American Indian or Alaska Native | 2 (0.3) | 0 (0) | | 2 (0.3) | 0 (0) | |
| Asian | 38 (6.5) | 4 (4.9) | | 39 (6.8) | 3 (3.3) | |
| Black, African American, or African | 37 (6.3) | 4 (4.9) | | 29 (5) | 12 (13) | |
| Hispanic, Latino, or Spanish | 42 (7.2) | 5 (6.1) | | 44 (7.6) | 3 (3.3) | |
| White | 425 (72.4) | 66 (80.5) | | 420 (72.8) | 71 (77.2) | |
| Multiple | 31 (5.3) | 2 (2.4) | | 31 (5.4) | 2 (2.2) | |
| Declined to answer | 9 (1.5) | 1 (1.2) | | 9 (1.6) | 1 (1.1) | |
| Other | 3 (0.5) | 0 (0) | | 3 (0.5) | 0 (0) | |
| At least one COVID-19 vaccine, n (%) | 555 (94.5) | 82 (100) | .15 | 548 (95) | 89 (96.7) | .65 |
| Preexisting conditions, n (%) | | | | | | |
| Asthma | 105 (17.9) | 12 (14.6) | >.99 | 104 (18) | 13 (14.1) | .59 |
| Autoimmune condition | 55 (9.4) | 8 (9.8) | >.99 | 57 (9.9) | 6 (6.5) | .59 |
| Cancer | 26 (4.4) | 9 (11) | .15 | 29 (5) | 6 (6.5) | .65 |
| Chronic bronchitis | 9 (1.5) | 1 (1.2) | >.99 | 10 (1.7) | 0 (0) | .59 |
| COPD[b] | 3 (0.5) | 0 (0) | >.99 | 2 (0.3) | 1 (1.1) | .59 |
| Diabetes | 36 (6.1) | 8 (9.8) | .65 | 35 (6.1) | 9 (9.8) | .59 |
| Emphysema | 1 (0.2) | 0 (0) | >.99 | 1 (0.2) | 0 (0) | >.99 |
| Heart disease | 16 (2.7) | 5 (6.1) | .59 | 20 (3.5) | 1 (1.1) | .59 |
| Heart failure | 2 (0.3) | 1 (1.2) | .65 | 2 (0.3) | 1 (1.1) | .59 |
| High blood pressure | 116 (19.8) | 16 (19.5) | >.99 | 111 (19.2) | 21 (22.8) | .63 |
| Other lung condition | 5 (0.9) | 2 (2.4) | .59 | 5 (0.9) | 2 (2.2) | .59 |
| None | 304 (51.8) | 36 (43.9) | .59 | 298 (51.6) | 42 (45.7) | .59 |
| Acute phase symptom surveys, mean (SD) | | | | | | |
| Systemic symptoms | 3.5 (2.3) | 3.9 (2.6) | .37 | 3.6 (2.4) | 2.7 (1.9) | .005 |
| Respiratory symptoms | 4.5 (2.4) | 5.6 (2.1) | .005 | 4.7 (2.5) | 4.2 (1.8) | .03 |
| Gastrointestinal symptoms | 1.0 (1.4) | 1.2 (1.7) | .37 | 1.1 (1.5) | 0.8 (1.0) | .03 |
| Neurological symptoms | 0.4 (1.1) | 0.3 (0.6) | .37 | 0.4 (1.1) | 0.3 (0.7) | .46 |
| Other symptoms | 0.3 (0.9) | 0.4 (1.1) | .51 | 0.4 (0.9) | 0.3 (0.9) | .83 |

[a]The acute phase symptom category measures are cumulative and represent the number of survey days across the 15-day acute study phase during which participants reported experiencing each symptom type (maximum 8 surveys).



XSL•FO
RenderX

[b]COPD: chronic obstructive pulmonary disease.

### *Symptom Frequencies*

Respiratory symptoms were the most commonly reported symptoms overall, with a mean number of 4.8 (SD 2.4) surveys in the NPR group and 4.5 (SD 2.4) surveys in the control group ($P$=.13; Table 1). The viral rebound group tended to report slightly more surveys with symptoms than those without viral rebound, with respiratory symptoms being the only statistically significant difference ($P$=.005; Table 2).

## Discussion

### Principal Findings

This COVID-19 Rebound Study found that participants who independently decided to take at least one dose of NPR had triple the 15-day cumulative incidence of viral test rebound and double the cumulative incidence of symptom rebound than the control group. These findings persisted in our secondary analysis after adjusting for baseline characteristics that may confound the association between NPR and COVID-19 rebound. Furthermore, when restricting the analysis to participants who achieved initial clearance before day 15, NPR use was significantly associated with a higher probability of both viral rebound (19.1% vs 7.2%; $P$<.001) and symptom rebound (47.7% vs 16.9%; $P$<.001) at study completion.

The estimates in our study are comparable to other studies, which have reported cumulative incidence of viral rebound in NPR treatment populations from 4% to 27%, and symptom rebound from 0.8% to 32% [21-23]. Studies with lower estimates tended to be secondary analyses of randomized controlled trials or retrospective analyses of electronic medical record data, which may only capture more severe cases [12,22-26]. Studies with higher estimates tended to use more sensitive measures of viral load, and in some cases, they did not require full resolution of symptoms for participants to be eligible for symptom rebound [21,27]. A strength of our study is that it was designed with rebound as the primary outcome and used frequent telehealth, proctor-verified home testing, and symptom surveys. Furthermore, in contrast to most other studies, we did not exclude participants who took partial doses of NPR, which was more common than taking the full 5-day prescription (n=226, 33.7% none, n=246, 36.7% partial, and n=197, 29.4% full), allowing for more accurate real-world estimates of rebound.

We found minimal overall differences in time to initial viral or symptom clearance between the NPR and control groups, suggesting that further investigation was needed into the discordance between how NPR reduces the potency of the virus to ultimately reduce the progression to severe COVID-19, as found in previous studies, and the identified increased rates of rebound or persistent virus. Time to initial symptom clearance was longer than viral clearance, consistent with another prospective study [21]. The tendency for symptoms to remain after testing negative was further seen in examinations of consistency between positive tests and symptoms across the 15-day acute phase follow-up (Table S6 in Multimedia Appendix 1). Of note, we captured self-reported presence or absence of symptoms, whereas symptom severity might yield additional insights.

An additional finding from our study is that at 5 days after enrollment (corresponding to up to 7 days after first turning positive), more than one-third (244/669, 36.5%) of participants remained positive on rapid antigen tests (Figure 3). Although Centers for Disease Control and Prevention guidance has largely recommended ending isolation after 5 days, our findings suggest that individuals may remain infectious for longer and should therefore remain cautious.

Numerous mechanisms have been hypothesized for why rebound occurs. The constant push and pull between virus and immunity drives “predator-prey dynamics,” where the virus’ presence triggers the activation of immune defenses, and virus clearance signals these immune defenses to shift toward a recovery mode [28]. If the shift is too early, the relaxed defenses can allow remaining virus to rapidly re-expand. Regarding NPR and increased rebound risk, it is possible that early treatment initiation pharmaceutically drives down virus growth while abrogating the need to raise a robust immunological defense. Thus, upon removal of NPR, remaining virus could replicate at a rate that outpaces immune defenses, leading to a virus rebound until defenses catch up. Consistent with many COVID-19 symptoms in an era of widespread immunity that represent activating immune defenses (ie, fever and congestion), rebound symptoms generally reflect mounting immune defenses. If immunity must increase to clear residual virus after completion of NPR treatment, many individuals would be expected to experience symptoms; this would lead to a more common symptom rebound phenomenon and a less common full virus rebound, which aligns with our results [29-31].

The COVID-19 Rebound Study was not designed to investigate the efficacy of NPR. Prior studies clearly demonstrate the impact of NPR on reducing COVID-19–related hospitalizations and severe outcomes [3-7,15]. On the basis of our rebound-related findings, our recommendation to clinical health care providers prescribing NPR is to inform patients of the higher likelihood of viral and symptom rebound, that there is insufficient evidence regarding the severity of these symptoms, and that the evidence supporting reduction in hospitalizations and progression to severe COVID-19 is strong.

This was a nonrandomized study that may be limited by selection bias, recall bias, and an unbalanced sample size. While NPR was the primary COVID-19 oral antiviral therapeutic, we did not systematically capture use of emerging alternative antiviral therapies in the control group, which may influence rebound estimates. Our definition of symptom rebound (≥1 symptom after resolution) is sensitive but may have lacked specificity, potentially overestimating clinically meaningful rebound events. Another limitation is that we required enrollment within 48-hours of a positive rapid antigen test; however, participants may have had symptoms and disease for longer durations prior to self-testing and/or asymptomatic participants may have tested at an unknown time point after infection, thereby leading to underestimates on time to viral or



XSL•FO
RenderX

symptom clearance. Timing estimates were also impacted by the length of the study as not all participants experienced clearance by day 15. The relatively large numbers of participants in both groups helps to mitigate, though not necessarily eliminate, the impact of these potential biases as they pertain to relative differences between groups. The social media recruitment arm was opened partway through the study primarily to increase the number of participants who were not intending to take NPR, as those recruited through the eMed platform may have been drawn to the test-to-treat virtual care paradigm. The analysis excluded 27% (248/917) of consented participants due to missing data that precluded capture of potential rebound. Excluded participants tended to be less vaccinated, NPR users, and non-White (Table S7 in Multimedia Appendix 1). There was also nonresponse or dropout during the study, whereby 67.6% (452/669) of the analysis sample completed all acute study phase surveys.

## Conclusions

In summary, this prospective observational study of COVID-19–positive outpatients showed that NPR treatment decision was associated with a substantially increased risk of both viral and symptom COVID-19 rebound. No difference was found in the overall time to initial clearance of viral test positivity or symptoms between the NPR and control groups, although the NPR group was older and had more preexisting conditions at baseline. With this important caveat for interpretation of our findings, there is a clear-cut need for more durable antiviral therapies or regimens and for improved knowledge regarding optimal patient selection.

## Acknowledgments

The authors thank Katie Quartuccio for planning and preparing participant materials for biosample collection and return, and Yuray Rodriquez for leading the kitting and logistics associated with this research.

## Funding

This work was supported by eMed, the National Institute of Allergy and Infectious Disease (grant 3U01AI151812-03S2), and the National Center for Advancing Translational Sciences (grant UM1TR004407).

## Data Availability

Data from the COVID-19 Rebound Study carry a risk of reidentification and therefore cannot be shared in accordance with the approved institutional review board protocol. Questions regarding access to the dataset should be directed to the corresponding authors.

## Authors' Contributions

GQ, DCC, EGS, JP, ER, LR, KGA, EJT, JAP, and MJM contributed to the conception and design of the study. JKK and KK performed the data analysis. JKK, MJM, and JAP interpreted the data and drafted the manuscript. DCC, JP, ER, KGA, and EJT provided substantial revisions. All authors reviewed and approved the final draft.

## Conflicts of Interest

MJM is chief scientific officer of eMed, a home test-to-treat company. He is the owner of BioHealth Engineering, a consulting firm providing advisory and consulting services to numerous diagnostics and life science companies. All other authors declare no other conflicts of interest.

Multimedia Appendix 1

The figures show time to initial viral and symptom clearance, time from clearance to rebound, and duration of rebound. The tables show differences between the two outreach cohorts, sensitivity analyses of 15-day COVID-19 cumulative incidence estimates, Cox proportional hazard regressions for symptom rebound, congruence of viral and symptom reports, and characteristics of participants included vs excluded from analyses.

[DOCX File, 2624 KB - i-jmr_v15i1e80263_app1.docx ]

## References

1. Surveillance and data analytics. US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/covid/php/surveillance/index.html [accessed 2024-04-08]
2. Coronavirus (COVID-19) update: FDA authorizes first oral antiviral for treatment of COVID-19. U.S. Food & Drug Administration. 2021. URL: https://web.archive.org/web/20211229172018/https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-first-oral-antiviral-treatment-covid-19 [accessed 2024-04-08]
3. FDA approves first oral antiviral for treatment of COVID-19 in adults. U.S. Food & Drug Administration. 2023. URL: https://tinyurl.com/2npxrtre [accessed 2023-05-25]



XSL•FO
RenderX

4. COVID-19 treatment clinical care for outpatients. U.S. Centers for Disease Control and Prevention. URL: https://tinyurl.com/yxjd5wxz [accessed 2024-04-08]
5. Living guidance for clinical management of COVID-19. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-clinical-2021-2 [accessed 2024-04-08]
6. Coronavirus disease 2019 (COVID-19) treatment guidelines. National Institutes of Health. 2024. URL: https://www.ncbi.nlm.nih.gov/books/NBK570371/pdf/Bookshelf_NBK570371.pdf [accessed 2024-04-29]
7. Hammond J, Leister-Tebbe H, Gardner A, et al. Oral nirmatrelvir for high-risk, nonhospitalized adults with Covid-19. N Engl J Med 2022 Apr 14;386(15):1397-1408. [doi: 10.1056/NEJMoa2118542] [Medline: 35172054]
8. Rubin R. Paxlovid is effective but underused-here's what the latest research says about rebound and more. JAMA 2024 Feb 20;331(7):548-551. [doi: 10.1001/jama.2023.28254] [Medline: 38294771]
9. Wilcock AD, Kissler S, Mehrotra A, et al. Clinical risk and outpatient therapy utilization for COVID-19 in the Medicare population. JAMA Health Forum 2024 Jan 5;5(1):e235044. [doi: 10.1001/jamahealthforum.2023.5044] [Medline: 38277170]
10. Appaneal HJ, LaPlante KL, Lopes VV, et al. Nirmatrelvir/ritonavir utilization for the treatment of non-hospitalized adults with COVID-19 in the national Veterans Affairs (VA) healthcare system. Infect Dis Ther 2024 Jan;13(1):155-172. [doi: 10.1007/s40121-023-00910-1] [Medline: 38217842]
11. Rosen A. More Americans could benefit from Paxlovid for COVID infection. Johns Hopkins Bloomberg School of Public Health. 2024. URL: https://publichealth.jhu.edu/2024/why-more-people-should-be-prescribed-paxlovid-for-covid [accessed 2026-05-03]
12. Harrington PR, Cong J, Troy SB, et al. Evaluation of SARS-CoV-2 RNA rebound after nirmatrelvir/ritonavir treatment in randomized, double-blind, placebo-controlled trials - United States and international sites, 2021-2022. MMWR Morb Mortal Wkly Rep 2023 Dec 22;72(51):1365-1370. [doi: 10.15585/mmwr.mm7251a2] [Medline: 38127674]
13. Callaway E. COVID rebound is surprisingly common - even without Paxlovid. Nature 2022 Aug 11. [doi: 10.1038/d41586-022-02121-z] [Medline: 35953572]
14. Pandit JA, Radin JM, Chiang DC, et al. The Coronavirus disease 2019 rebound study: a prospective cohort study to evaluate viral and symptom rebound differences in participants treated with nirmatrelvir plus ritonavir versus untreated controls. Clin Infect Dis 2023 Jul 5;77(1):25-31. [doi: 10.1093/cid/ciad102] [Medline: 36810665]
15. Tian H, Yang C, Song T, et al. Efficacy and safety of Paxlovid (nirmatrelvir/ritonavir) in the treatment of COVID-19: an updated meta-analysis and trial sequential analysis. Rev Med Virol 2023 Sep;33(5):e2473. [doi: 10.1002/rmv.2473] [Medline: 37485774]
16. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
17. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
18. Pollard TJ, Johnson AE, Raffa JD, Mark RG. tableone: an open source Python package for producing summary statistics for research papers. JAMIA Open 2018;1(1):26-31. [doi: 10.1093/jamiaopen/ooy012] [Medline: 31984317]
19. Davidson-Pilon C. lifelines, survival analysis in Python. Zenodo. 2023. URL: https://zenodo.org/records/10456828 [accessed 2026-05-03]
20. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]
21. Smith-Jeffcoat SE, Biddle JE, Talbot HK, et al. Symptoms, viral loads, and rebound among COVID-19 outpatients treated with nirmatrelvir/ritonavir compared with propensity score-matched untreated individuals. Clin Infect Dis 2024 May 15;78(5):1175-1184. [doi: 10.1093/cid/ciad696] [Medline: 37963102]
22. Hammond J, Fountaine RJ, Yunis C, et al. Nirmatrelvir for vaccinated or unvaccinated adult outpatients with Covid-19. N Engl J Med 2024 Apr 4;390(13):1186-1195. [doi: 10.1056/NEJMoa2309003] [Medline: 38598573]
23. Ranganath N, O'Horo JC, Challener DW, et al. Rebound phenomenon after nirmatrelvir/ritonavir treatment of Coronavirus disease 2019 (COVID-19) in high-risk persons. Clin Infect Dis 2023 Feb 8;76(3):e537-e539. [doi: 10.1093/cid/ciac481] [Medline: 35698452]
24. Petrakis V, Rafailidis P, Trypsianis G, Papazoglou D, Panagopoulos P. The antiviral effect of nirmatrelvir/ritonavir during COVID-19 pandemic real-world data. Viruses 2023 Apr 16;15(4):976. [doi: 10.3390/v15040976] [Medline: 37112956]
25. Wang L, Volkow ND, Davis PB, Berger NA, Kaelber DC, Xu R. COVID-19 rebound after Paxlovid treatment during Omicron BA.5 vs BA.2.12.1 subvariant predominance period. medRxiv. 2022 Aug 6 p. 2022.08.04.22278450. [doi: 10.1101/2022.08.04.22278450] [Medline: 35982673]
26. Camp D, Caputo M, Echevarria FM, Achenbach CJ. Clinical rebound after treatment with nirmatrelvir/ritonavir in COVID-19. BMC Infect Dis 2024 Sep 12;24(1):963. [doi: 10.1186/s12879-024-09842-8] [Medline: 39266964]
27. Edelstein GE, Boucau J, Uddin R, et al. SARS-CoV-2 virologic rebound with nirmatrelvir-ritonavir therapy: an observational study. Ann Intern Med 2023 Dec;176(12):1577-1585. [doi: 10.7326/M23-1756] [Medline: 37956428]




XSL•FO
RenderX

28. Pilyugin SS, Antia R. Modeling immune responses with handling time. Bull Math Biol 2000 Sep;62(5):869-890. [doi: 10.1006/bulm.2000.0181] [Medline: 11016088]
29. Phan T, Ribeiro RM, Edelstein GE, et al. Modeling suggests SARS-CoV-2 rebound after nirmatrelvir-ritonavir treatment is driven by target cell preservation coupled with incomplete viral clearance. J Virol 2025;99(3):e0162324. [doi: 10.1128/jvi.01623-24] [Medline: 39902924]
30. Esmaeili S, Owens K, Wagoner J, Polyak SJ, White JM, Schiffer JT. A unifying model to explain frequent SARS-CoV-2 rebound after nirmatrelvir treatment and limited prophylactic efficacy. Nat Commun 2024 Jun 28;15(1):5478. [doi: 10.1038/s41467-024-49458-9] [Medline: 38942778]
31. Perelson AS, Ribeiro RM, Phan T. An explanation for SARS-CoV-2 rebound after Paxlovid treatment. medRxiv. 2023 Jun 1 p. 2023.05.30.23290747. [doi: 10.1101/2023.05.30.23290747] [Medline: 37398088]

## Abbreviations

**CLIA:** Clinical Laboratory Improvement Amendments
**NPR:** nirmatrelvir plus ritonavir
**REDCap:** Research Electronic Data Capture
**SRTI:** Scripps Research Translational Institute
**STROBE:** Strengthening the Reporting of Observational Studies in Epidemiology

*Edited by M Balcarras; submitted 21.Jul.2025; peer-reviewed by A Girma, M Cruciani; revised version received 31.Mar.2026; accepted 31.Mar.2026; published 02.Jun.2026.*

*Please cite as:*
*Kueper JK, Kottilil K, Quer G, Chiang DC, Spencer EG, Purushotham J, Ramos E, Roumani L, Andersen KG, Topol EJ, Pandit JA, Mina MJ*
*COVID-19 Rebound in Nirmatrelvir Plus Ritonavir Treatment and Control Groups: Prospective Cohort Study*
*Interact J Med Res 2026;15:e80263*
*URL: https://www.i-jmr.org/2026/1/e80263*
*doi:10.2196/80263*

© Jacqueline K Kueper, Kalyani Kottilil, Giorgio Quer, Danielle C Chiang, Emily G Spencer, Jyothi Purushotham, Edward Ramos, Leila Roumani, Kristian G Andersen, Eric J Topol, Jay A Pandit, Michael J Mina. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 2.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Factors Affecting the Accuracy of Clinical Coding in the Casemix System in a Teaching Hospital: Cross-Sectional Study

Amirah Azzeri[1], MBBS, MPH, PhD; Wan Mastura Wan Musaludin[2], MD, MPH; Hafiz Jaafar[1], MBBS, MPH, DrPH; Nurnabihah MD Hafidz[2], MD, MPH; Azimatun Noor Aizuddin[2], MD, MCommMed, PhD

[1]Public Health Unit, Department of Primary Health Care, Faculty of Medicine and Health Ssciences, Universiti Sains Islam Malaysia, Persiaran Ilmu, Bandar Baru Nilai, Nilai, Negeri Sembilan, Malaysia

[2]Department of Public Health, Faculty of Medicine, Universiti Kebangsaan Malaysia, Jalan Yaacob Latif, Bandar Tun Razak, Cheras, Kuala Lumpur, Malaysia

**Corresponding Author:**
Azimatun Noor Aizuddin, MD, MCommMed, PhD
Department of Public Health, Faculty of Medicine, Universiti Kebangsaan Malaysia, Jalan Yaacob Latif, Bandar Tun Razak, Cheras, Kuala Lumpur, Malaysia

## Abstract

**Background:** Accurate clinical coding is critical in the casemix system to ensure proper resource allocation, health care policy planning, and data reliability. Inaccurate coding can result in significant financial losses to the hospital. To date, the implications of inaccurate coding in casemix implementations in Malaysia have rarely been explored.

**Objective:** This study aimed to evaluate the accuracy of clinical coding and identify factors associated with accurate coding practices in a major teaching hospital in Malaysia.

**Methods:** A cross-sectional study was conducted using 445 inpatient discharge records from Hospital Canselor Tuanku Muhriz from January 2023 to December 2023. Stratified random sampling was applied across 4 departments. Coding accuracy was determined by comparing electronic medical record entries to a gold-standard set by trained coders and specialists. Descriptive and bivariate analyses were performed accordingly.

**Results:** This study found that the overall clinical coding accuracy was 76.4% (340/445). Documentation completeness ($P$<.001) and coding turnaround time ($P$=.02) were significantly associated with coding accuracy. Other variables such as patients' sex, patient age, department, and coder experience were not significantly associated. Clinical coding accuracy in this setting was suboptimal.

**Conclusions:** Continuous training and regular audits are recommended to improve coding quality and ensure reliable casemix data for policy and funding decisions.

***(Interact J Med Res 2026;15:e93769)*** doi:10.2196/93769

**KEYWORDS**

clinical coding; casemix; diagnosis-related group; ICD-10; coding errors; coding accuracy

## Introduction

Clinical coding serves as the backbone of the casemix system, translating detailed patient diagnoses and procedures into standardized alphanumeric codes, such as those defined by the *International Classification of Diseases* and related classifications [1,2]. These codes are crucial for enabling workload analysis, hospital performance benchmarking, resource allocation, reimbursement processes, epidemiological research, and national health planning [2,3]. In Malaysia, the casemix system, specifically the Malaysian Diagnosis-Related Group (My-DRG), was first introduced at a major university hospital under the Ministry of Higher Education, aiming to improve efficiency and accountability in health care service delivery [4].

Despite the well-established role of casemix systems internationally, Malaysia's nationwide implementation continues to face challenges, including shortages of trained coding professionals, inadequate infrastructure, and limited political and institutional commitment. One of the critical barriers to broader adoption is persistent concern over coding accuracy. Coding errors can lead to significant negative consequences such as financial discrepancies, misallocation of health care resources, flawed health policy decisions, and compromised data quality for research and planning [5,6]. Accurate clinical coding is especially vital in teaching hospitals, where patient caseloads are typically complex and involve multidisciplinary management.

The process of clinical coding is inherently intricate, involving the translation of complex diagnostic and procedural data into



XSL•FO
RenderX

standardized codes while ensuring fidelity to clinical records. This process depends heavily on accurate and comprehensive documentation by health care providers, coders' competence, and institutional systems supporting coding practices [7]. Studies have consistently demonstrated that coding errors arise from various sources, including incomplete or ambiguous medical documentation, coder experience and training gaps, and systemic factors such as departmental practices or weak organizational culture regarding data quality [7-9].

Several patient-related factors may influence coding accuracy. Older patients, for instance, often present with complex medical conditions and multiple comorbidities, making precise coding more difficult [10,11]. Similarly, longer hospital stays may indicate complex cases requiring multifaceted care, which can increase the potential for coding errors, although findings in the literature on this relationship remain mixed [12,13]. Departmental factors also play a role; surgical departments typically exhibit higher coding accuracy due to standardized procedural documentation, whereas medical and obstetric units face greater challenges due to diagnostic complexity and unplanned interventions [7,13].

Documentation quality is a critical determinant. Poorly written records, nonstandard abbreviations, or incomplete discharge summaries have all been linked to higher coding error rates [4,8]. Electronic medical records (EMRs), while intended to enhance documentation, can sometimes contribute to inaccuracies due to poor usability or reliance on default entries and copying-and-pasting practices [14,15]. Coder-related characteristics, including age, educational level, experience, and ongoing training, have also been shown to influence accuracy. Coders with higher qualifications and those who receive regular training tend to perform better [16-18]. In contrast, high workloads and time pressure are consistently associated with decreased accuracy [9,19].

Although international literature highlights these factors, research within the Malaysian context, particularly in teaching hospitals, is limited. Available studies suggest that coding accuracy levels in Malaysia vary widely, with error rates ranging from 25% to nearly 50% in different settings [6,7]. This variability underscores the need for localized investigations to identify modifiable factors that can enhance coding practices and data quality.

Therefore, this study aimed to determine the level of clinical coding accuracy and explore the patient, coder, documentation, and systemic factors associated with coding accuracy among discharged inpatients in a Malaysian teaching hospital. The findings will provide critical evidence to inform interventions aimed at improving coding practices, supporting effective casemix implementation, and strengthening the health system's capacity for data-driven decision-making.

## Methods

### Overview

A cross-sectional study was conducted at Hospital Canselor Tuanku Muhriz (HCTM), a university hospital in Malaysia. It was formerly known as Universiti Kebangsaan Malaysia Medical Centre. It serves as a referral center for patients from Klang Valley and is also the first hospital in Malaysia that implemented a casemix system. HCTM requires coders to conduct active and ongoing diagnosis and procedure coding exercises.

This was a cross-sectional comparative study using secondary data obtained from inpatient discharge lists between January 2023 and December 2023. The aim was to assess the accuracy of clinical coding for the principal diagnoses. All patient information for the episode of admission was stored in the EMR system in the hospital database. In addition, the casemix unit at the International Centre for Casemix and Clinical Coding (ITCC) HCTM performed the diagnosis and procedure codings by extracting data from the EMR system. The database was accessed from June 1, 2024, to July 31, 2024. The authors had no access to information that could identify individual patients during or after data collection. Stratified random sampling was applied to obtain patients from 4 groups of departments: medicine, surgery, orthopedics, and obstetrics and gynecology. Subsequently, systematic random sampling was used via a computer system to select patient records from each department to ensure equivalent representativeness. The required sample size was calculated using the formula for finite populations, targeting a 95% confidence level and 5% margin of error. An estimated population of 31,103 discharges in 2023 yielded a required sample size of 384. A 20% buffer for incomplete data was added, yielding a final target of 445 records.

This study included secondary information from the discharge summaries in the EMR system, casemix-coded patient data, and coder demographic profiles. A comparison was made between the coders' codes and the codes reassigned by a trained medical officer and casemix experts. The gold-standard codes were determined through a structured review process conducted by a trained medical officer with casemix coding experience and verified by certified casemix experts at HCTM. Each sampled record was independently reviewed against the discharge summary in the EMR, and the reassigned code was agreed upon through consensus between the medical officer and the casemix expert. In cases of disagreement, a third senior casemix expert served as the final arbiter to ensure a consistent and predefined adjudication process. To minimize assessment bias, the trained medical officer and casemix experts who reassigned the principal diagnosis codes were blinded to the original coder-assigned codes at the time of review. The original codes were only compared after the gold-standard codes had been independently assigned, ensuring that the reassignment process was not influenced by knowledge of the coders' prior decisions. In this study, the dependent variable was coding accuracy, defined as whether the principal diagnosis was coded correctly by the casemix coder, identical to the codes assigned by the trained medical officer or casemix experts. The data were categorized into 2 groups: accurate or inaccurate. This study deliberately focused on the accuracy of the principal diagnosis code only rather than secondary diagnoses or procedure codes. This decision was based on the fact that the principal diagnosis is the primary driver of My-DRG casemix grouping and hospital reimbursement in the Malaysian casemix system, making it the most consequential code for financial and policy outcomes. It



XSL•FO
RenderX

is important to note that this study evaluated the accuracy of the *International Classification of Diseases, Tenth Revision*, principal diagnosis coding specifically rather than the resulting My-DRG group assignment itself. My-DRG grouping accuracy depends on the joint accuracy of multiple coded elements (principal diagnosis, secondary diagnoses, procedures, age, and length of stay) and was beyond the scope of this study. Furthermore, restricting the analysis to the principal diagnosis allowed for a focused and interpretable measure of coding performance, consistent with several prior studies in similar health care settings [7].

Four components of independent variables were included: (1) patient-related factors such as age, sex, and length of stay; (2) clinical departments; (3) documentation factor; and (4) coder-related factors, namely, length of service, educational level, and time taken to assign a clinical code. The length of stay in hospital was defined as the number of days that the patient stayed in the hospital and was further divided into inliers and outliers. The discharge summary was reviewed for completeness of diagnosis. Documentation was considered complete if an actual primary diagnosis was recorded. Specifically, a discharge summary was considered complete if it contained an explicitly stated primary (principal) diagnosis written in full clinical terminology without reliance on abbreviations, symbols, or ambiguous shorthand. Summaries that contained only nonspecific descriptors (eg, “unwell,” “fever,” or “query diagnosis”), left the primary diagnosis field blank, or contained illegible entries were classified as incomplete. This operational definition was consistently applied by the reviewing medical officer across all sampled records.

For coder-related factors, the length of service was divided into less than 10 years and greater than or equal to 10 years. This threshold was used as it is consistent with the minimum reckonable service period for confirmed or pensionable status and optional retirement eligibility under Malaysia’s Pensions Act 1980 (Act 227), section 12 [20]. Educational level was categorized as having a degree or not, and the time taken for the coding officer to complete coding was calculated from the date of discharge until the date when the coding data were finalized. This period was divided into 2 categories: 3 days and below and 4 days and above. These cutoff points are consistent with the international discharged, not final coded benchmark used in health information management departments, where 3 days after discharge is the standard threshold for flagging a record as overdue for coding completion [21].

Descriptive analysis was conducted to present the background and study variables. Bivariate analyses were performed to understand the association between independent variables and the dependent variable. All associations were considered statistically significant if the *P* value was below .05. All data were recorded in Microsoft Excel and analyzed using SPSS (version 29; IBM Corp). Bivariate analysis using the Pearson chi-square test was performed to examine whether the coding accuracy status (accurate vs inaccurate) differed significantly across subgroups of each patient and coder characteristic. This approach was adopted to identify which specific variables were statistically associated with coding inaccuracy, thereby allowing for the prioritization of variables for targeted quality improvement interventions. In addition to the test of association, the coding accuracy rate for each variable category was calculated as the proportion of accurately coded records out of the total records in that category to provide a more interpretable measure of coding performance at the variable level. All 445 sampled records had complete data for discharge summary completeness status and coder-related variables. No records required exclusion due to missing data.

### Ethical Considerations

This study received ethics approval from the Medical Research and Ethics Committee of Universiti Kebangsaan Malaysia, with approval reference number JEP-2024-289. All patient data were anonymized to ensure participant confidentiality, and no individual participants could be identified during or after data collection. Informed consent was not required as the study used deidentified retrospective administrative data.

## Results

Of 445 inpatient records, this study found that 340 (76.4%) were coded correctly, whereas 105 (23.6%) were coded incorrectly. Of the 445 records included, 174 (39.1%) were of male patients, and 271 (60.9%) were of female patients. Regarding patient age range, the highest frequency was 30 to 39 years at 21.1% (n=94), followed by 60 to 69 years at 18.2% (n=81). The lowest frequency was in the ranges of 0 to 9 years and 10 to 19 years at 4.7% (n=21) each.

It was also found that 76.6% (341/445) of the inpatient records were categorized as inliers and 23.4% (104/445) were categorized as outliers. Among the 445 samples, the largest number corresponded to the medical department (n=123, 27.6%), followed by the obstetrics and gynecology department (n=110, 24.7%) and both the surgery and orthopedics departments (n=106, 23.8% each). Regarding the documentation of discharge summaries, 88.1% (392/445) of the records were complete, and 11.9% (53/445) were incomplete. Coders without a degree coded 28.1% (125/445) of the records, and those with a degree coded 71.9% (320/445). The number of records coded by coders with less or more than 10 years of service was almost equal. In terms of coding turnaround time, 72.1% (321/445) of the records were coded within 3 days or less, and 27.9% (124/445) were coded in 4 days or more. The descriptive analysis of patient sociodemographic and clinical characteristics and coder profiles is shown in Table 1.



XSL•FO
RenderX

**Table .** Descriptive analysis of patient sociodemographic and clinical characteristics and coder profiles (N=445).

| Factors | Frequency, n (%) |
|---|---|
| Sex | |
| Male | 174 (39.1) |
| Female | 271 (60.9) |
| Age (y) | |
| 0-9 | 21 (4.7) |
| 10-19 | 21 (4.7) |
| 20-29 | 50 (11.2) |
| 30-39 | 94 (21.1) |
| 40-49 | 50 (11.2) |
| 50-59 | 38 (8.5) |
| 60-69 | 81 (18.2) |
| 70-79 | 59 (13.3) |
| ≥80 | 31 (7) |
| Length of stay | |
| Inliers | 341 (76.6) |
| Outliers | 104 (23.4) |
| Department | |
| Medical | 123 (27.6) |
| Surgical | 106 (23.8) |
| Orthopedic | 106 (23.8) |
| Obstetrics and gynecology | 110 (24.7) |
| Coders' educational level | |
| No degree | 125 (28.1) |
| Degree and higher | 320 (71.9) |
| Coders' length of service (y) | |
| <10 | 227 (51) |
| >10 | 218 (49) |
| Time taken for coding (d) | |
| ≤3 | 321 (72.1) |
| ≥4 | 124 (27.9) |

A total of 4 casemix coders were responsible for the 445 coded records included in this study. On average, each of them coded 100 to 110 cases. This uneven distribution reflects the actual workload allocation within the casemix unit at HCTM during the study period.

To determine which patient and coder characteristics were associated with coding accuracy, bivariate analyses were conducted alongside the calculation of accuracy rates per variable category. This dual approach allowed for identification of both the magnitude of coding accuracy for each subgroup and whether observed differences were statistically significant.

Table 2 shows the coding accuracy rate for each variable. Overall, 76.4% (340/445) of the records were coded accurately for the main diagnosis. At the variable level, coding accuracy varied across patient and coder characteristics. For sex, the accuracy rate was 79.9% (139/174) for male patients and 74.2% (201/271) for female patients, with no statistically significant association ($\chi^2(1)=1.9$; $P=.17$). Across age groups, accuracy ranged from 61.9% (13/21) in the group of 10 to 19 years to 84% (42/50) in the group of 40 to 49 years, although this variation was not statistically significant ($\chi^2(8)=8.8$; $P=.36$). For length of stay, accuracy was comparable between inliers (259/341, 76%) and outliers (81/104, 77.9%), with no significant difference ($\chi^2(1)=0.2$; $P=.69$).



XSL•FO
RenderX

**Table .** Coding accuracy by variable and association with accuracy status (N=445).

| Factors | Accurate (n=340), n (%) | Inaccurate (n=105), n (%) | Chi-square (*df*) | *P* value |
|---|---|---|---|---|
| Sex | | | 1.9 (1) | .17 |
| Male | 139 (40.9) | 35 (33.3) | | |
| Female | 201 (59.1) | 70 (66.7) | | |
| Age (y) | | | 8.8 (8) | .36 |
| 0 - 9 | 15 (4.4) | 6 (5.7) | | |
| 10 - 19 | 13 (3.8) | 8 (7.6) | | |
| 20 - 29 | 39 (11.5) | 11 (10.5) | | |
| 30 - 39 | 63 (18.5) | 31 (29.5) | | |
| 40 - 49 | 42 (12.4) | 8 (7.6) | | |
| 50 - 59 | 28 (8.2) | 10 (9.5) | | |
| 60 - 69 | 66 (19.4) | 15 (14.3) | | |
| 70 - 79 | 47 (13.8) | 12 (11.4) | | |
| ≥80 | 25 (7.4) | 6 (5.7) | | |
| Length of stay | | | 0.2 (1) | .69 |
| Inliers | 259 (76.2) | 82 (78.1) | | |
| Outliers | 81 (23.8) | 23 (21.9) | | |
| Department | | | 3.6 (3) | .31 |
| Medical | 98 (28.8) | 25 (23.8) | | |
| Surgical | 85 (25) | 21 (20) | | |
| Orthopedic | 79 (23.2) | 27 (25.7) | | |
| Obstetrics and gynecology | 78 (22.9) | 32 (30.5) | | |
| Discharge summary | | | 42.4 (1) | <.001 |
| Complete | 319 (93.8) | 73 (69.5) | | |
| Not complete | 21 (6.2) | 32 (30.5) | | |
| Coders' educational level | | | 0.1 (1) | .71 |
| No degree | 94 (27.6) | 31 (29.5) | | |
| Degree and above | 246 (72.4) | 74 (70.5) | | |
| Coders' length of service (y) | | | 1.5 (1) | .23 |
| <10 | 168 (49.4) | 59 (56.2) | | |
| >10 | 172 (50.6) | 46 (43.8) | | |
| Time taken for coding (d) | | | 5.9 (1) | .02 |
| ≤3 | 255 (75) | 66 (62.9) | | |
| ≥4 | 85 (25) | 39 (37.1) | | |

Regarding clinical department, the highest coding accuracy was observed in the surgical (85/106, 80.2%) and medical (98/123, 79.7%) departments, whereas obstetrics and gynecology recorded the lowest accuracy rate at 70.9% (78/110), followed by orthopedics at 74.5% (79/106). However, these differences were not statistically significant ($\chi^2(3)=3.6$; $P=.31$).

The completeness of discharge summaries showed the most striking difference in accuracy rates. Records with complete discharge summaries had an accuracy rate of 81.4% (319/392) compared to only 40.4% (21/52) for those with incomplete summaries, a difference that was highly statistically significant ($\chi^2(1)=42.4$; $P<.001$). For coder-related factors, accuracy rates were similar between coders without a degree (94/125, 75.2%) and those with a degree or a higher educational level (246/320, 76.9%), with no significant association ($\chi^2(1)=0.1$; $P=.71$). Similarly, coders with less than 10 years of service had an accuracy rate of 74.0% (168/227) compared to 78.9% (172/218) for those with 10 or more years, which was not statistically significant ($\chi^2(1)=1.5$; $P=.23$). Finally, records coded within 3



XSL•FO
RenderX

days of discharge had a higher accuracy rate of 79.4% (255/321) compared to 68.5% (85/124) for those coded after 4 or more days, a difference that was statistically significant ($\chi^2(1)=5.9$; $P=.02$).

Overall, demographic and departmental factors did not significantly influence coding accuracy, whereas documentation completeness and timely coding emerged as key contributors. These findings highlight the importance of structured documentation practices and effective coder workflow management.

## Discussion

### Principal Findings

This study aimed to assess the accuracy of clinical coding for principal diagnoses at a major Malaysian teaching hospital. The findings revealed an overall coding accuracy rate of 76.4% (340/445). These results are consistent with those of prior literature but highlight room for improvement. For example, Olagundoye et al [22] reported a slightly higher accuracy rate of 78.7% at Lagos State University Teaching Hospital. Worldwide, clinical coding accuracy varies substantially, with a systematic review by Burns et al [23] identifying a median accuracy rate of 83.2%, ranging from as low as 50.5% to as high as 97.8%. Our findings fall within this range but remain below the optimal threshold for high-performing health systems.

Interestingly, our error rate was lower than the 49.8% reported by Zafirah et al [7] and comparable to the 25.2% observed by Saizan et al [6]. This difference may be attributed to methodological differences: while our study focused solely on the principal diagnosis, many others have evaluated multiple diagnoses and procedural codes, which are inherently more complex and prone to error. Such methodological clarity is a strength of our study, allowing for a focused and interpretable measure of accuracy that reflects core hospital performance in coding.

One of the most significant findings was the strong association between the completeness of discharge summaries and coding accuracy. In our sample, 88.1% (392/445) of the records were complete, substantially higher than the 7% completeness rate reported by Zafirah et al [7]. This likely stems from our focus on the principal diagnosis, whereas previous studies have assessed full diagnostic and procedural completeness. Nevertheless, our statistical analysis confirmed that complete documentation significantly improves coding accuracy ($P<.001$). These results align with those of Taiwo Adeleke et al [24], who reported that incomplete or poor-quality documentation was a leading contributor to coding inaccuracy in Nigerian hospitals. Moreover, Foster et al [5] emphasize that continuous auditing and feedback, combined with electronic discharge summary systems, can meaningfully enhance both documentation and coding standards. Active clinician engagement in the documentation process, as advocated by Nouraei et al [25], is also vital as the absence of clear diagnostic notes can critically impair the coder's ability to assign appropriate codes.

Another notable finding was the statistically significant association between coding timeliness and accuracy. Records coded within 3 days of discharge were significantly more accurate than those coded after 4 or more days (75% vs 25%; $P=.02$). This supports the view that shorter coding turnaround times may reflect more efficient workflows or more straightforward cases. However, longer durations could also indicate increased care and attention to complex cases. Studies have found that coders who took additional time often achieved better accuracy, particularly for complex cases [26,27]. This raises the possibility that longer turnaround time may, in part, reflect greater case complexity rather than inefficient workflow alone, a distinction this study could not disentangle as case-level complexity (eg, number of comorbidities, diagnostic ambiguity, and length and diversity of documentation) was not independently measured. Coding complexity has been recognized elsewhere as a substantive, measurable construct that affects coder workload and throughput independent of coder efficiency [26,27], and future studies should consider incorporating a complexity measure (eg, number of secondary diagnoses, comorbidity index, or casemix index) to formally test whether case complexity confounds the turnaround time–accuracy relationship. Conversely, studies have also reported that prolonged delays might signal coder fatigue, systemic inefficiencies, or excessive workloads, factors that could compromise coding quality [28,29]. These findings suggest that turnaround time is a multifaceted indicator that could be used by hospital administrators to identify bottlenecks and improve workflow efficiency.

While the analysis of department-specific trends showed that obstetrics and gynecology had the highest rate of coding errors followed by orthopedics, internal medicine, and surgery, these differences were not statistically significant. Nevertheless, this trend aligns with findings by Zafirah et al [7] indicating that certain specialties may be more vulnerable to documentation and coding errors. This calls for department-specific interventions such as targeted coder training or closer clinician-coder collaboration, particularly in departments with complex case profiles.

Interestingly, coder-related factors such as educational attainment and years of experience did not significantly affect coding accuracy in this study. While this contrasts with literature suggesting that formal qualifications and experience enhance coder performance, our findings may be due to the small number of coders (n=4) and the relatively uniform training they receive at the ITCC. Future studies with larger and more diverse coder samples could provide more definitive insights into these relationships.

This study has several strengths. It is one of the few Malaysian investigations to focus specifically on principal diagnosis coding, providing a clear and interpretable metric of performance. It also combines administrative data with manually validated recoding, enhancing data reliability. Moreover, the statistically significant findings on discharge summary completeness and coding timeliness offer actionable insights for quality improvement.

However, some limitations must be acknowledged. This study was conducted at a single teaching hospital, limiting generalizability to other health care settings, including private



XSL•FO
RenderX

or nonacademic hospitals. Teaching hospitals such as HCTM typically have access to specialist input; structured coder training programs; and more complex, tertiary-level case mixes than nonteaching public hospitals. Nevertheless, teaching hospitals may also differ from private hospitals in EMR infrastructure, documentation culture, and staffing ratios of coders to caseload. The coding accuracy patterns observed in this study, particularly the influence of documentation completeness and turnaround time, may therefore not directly generalize to district or state public hospitals or private health care settings, and multisite studies are recommended to confirm these findings across different hospital types.

Additionally, the small number of coders limited our ability to explore demographic factors robustly. With only 4 coders contributing to the 445 coded records, coder-level variables (such as educational level and years of service) were effectively compared across just 4 clusters rather than 445 independent units, substantially limiting statistical power to detect true coder-level effects. Nonsignificant associations for these variables should therefore be interpreted with caution rather than as confirmed null findings. This study also did not engage clinicians directly to understand root causes of documentation lapses, which could have enriched the analysis. Additionally, a formal interrater reliability assessment between the trained medical officer and casemix experts was not conducted, which is acknowledged as a limitation of this study. Future studies should incorporate a structured reliability assessment such as the Cohen κ to strengthen the validity of the gold-standard coding process.

It is acknowledged that the use of separate bivariate chi-square tests to examine associations between individual variables and coding accuracy status has limitations as this approach does not account for confounding between variables. In particular, potential confounders such as comorbidity burden, admission type (elective vs emergency), discharge disposition, and overall casemix severity were not captured in this dataset and could not be adjusted for. As this was a preliminary descriptive study designed to characterize bivariate associations rather than isolate independent predictors, these factors represent important variables for inclusion in future multivariable analyses. Future studies should consider using multivariable logistic regression, with coding accuracy status as the dependent variable and coder-related factors, including educational level, years of service, time taken for coding, and department, as independent variables. Such an approach would yield adjusted estimates and provide more robust evidence on the independent predictors of coding inaccuracy. In this analysis, chi-square tests were used to examine bivariate associations between coding accuracy and the variables of interest, rather than to identify independent predictors. The findings were intended to describe patterns of association and inform variables that may warrant further investigation in future multivariable analyses.

The analysis was limited to bivariate associations as an exploratory first step. A multivariable logistic regression model was not developed as this study aimed to describe preliminary associations rather than establish causal or predictive relationships. The small number of coders (n=4) further precluded robust multivariable modeling, and future studies with larger coder samples are encouraged to build on these findings. A multivariable logistic regression model was therefore not pursued at this stage as the primary aim was to describe the pattern of associations across variables rather than determine independent predictors. Future studies with larger and more diverse samples are recommended to perform logistic regression analysis to confirm these preliminary findings and identify independent predictors of coding inaccuracy.

Additionally, a sensitivity analysis restricting the sample to records with complete discharge summaries was not performed in this preliminary study. Given that documentation completeness was the strongest predictor of accuracy, future work should also examine whether other associations (eg, turnaround time and years of service) persist when analysis is restricted to records with complete documentation to rule out confounding by documentation completeness itself.

Despite these limitations, the findings have important implications for policymakers and hospital administrators. Improving the completeness and clarity of discharge summaries through training, digital templates, and clinical documentation improvement initiatives can directly enhance coding accuracy. In turn, more accurate coding enables better casemix classification, fairer hospital reimbursements, and more reliable health service planning. The significant relationship between coding timeliness and accuracy also suggests the need for balanced workload distribution and process optimization to reduce coding delays without compromising quality.

Hospitals should implement a structured, recurring coding audit program, for example, auditing 3.5% to 5% of monthly coded records on a weekly cycle, consistent with established health information management auditing practice, rather than relying on ad hoc or annual reviews [30]. Coder feedback should be delivered promptly after each audit cycle, with individualized scorecards highlighting recurring error patterns (eg, specific diagnosis categories or documentation gaps). A formal, structured clinician query process following established frameworks such as the American Health Information Management Association and Association of Clinical Documentation Integrity Specialists *Guidelines for Achieving a Compliant Query Practice* should be adopted to standardize how coders seek clarification from clinicians when discharge documentation is ambiguous or incomplete [30]. Standardized discharge summary templates with mandatory fields for principal diagnosis and key clinical indicators should also be piloted to reduce documentation-related coding errors, the single strongest predictor of inaccuracy identified in this study.

As health systems move toward performance-based financing and data-driven policy, coding accuracy becomes not merely a technical concern but also a strategic priority. In this context, this study's findings offer a timely reminder that investments in documentation quality, coder support, and integrated clinical workflows are essential to realizing the full value of casemix systems in Malaysia and beyond.

## Conclusions

The results of this study showed that clinical coding accuracy was 76.4% (340/445), lower than the key performance index



XSL•FO
RenderX

set by the Ministry of Health of Malaysia, which is 90%. The findings underscore the importance of complete documentation and timely coding in enhancing the accuracy of clinical coding in casemix systems. Hospitals should invest in ongoing coder training and implement systematic audits to ensure high-quality data, which is critical for planning, financing, and evaluating health care services.

## Acknowledgments

The authors are grateful to all officers at the Health Informatics Centre, Hospital Canselor Tuanku Muhriz, Universiti Kebangsaan Malaysia. They confirm that generative AI tools were used solely for English-language editing of this manuscript. No AI was used for data generation, analysis, or scientific interpretation.

## Funding

The authors declared no financial support was received for this work.

## Authors' Contributions

AA and WMWM wrote the manuscript. WMWM, HJ, and NMDH conducted all statistical analyses. ANA provided supervision throughout the study and the writing of the manuscript. All authors reviewed the final manuscript.

## Conflicts of Interest

None declared.

## References

1. Anian S, Ahmad Nizaruddin M, Subramaniam G, Marimuthu J, Sujak S. Perception of nurses on ease of use, usefulness and system quality of electronic clinical documentation (ECD) at private hospitals in Klang Valley, Malaysia. IeJSME 2021;15(2):5-18. [doi: 10.56026/imu.15.2.5]
2. López-García G, Jerez JM, Ribelles N, Alba E, Veredas FJ. Explainable clinical coding with in-domain adapted transformers. J Biomed Inform 2023 Mar;139:104323. [doi: 10.1016/j.jbi.2023.104323] [Medline: 36813154]
3. Garis panduan audit dokumentasi klinikal dan kod klasifikasi [Article in Malay]. Ministry of Health Malaysia. 2025. URL: https://www.moh.gov.my/images/04-penerbitan/garis-panduan-kkm/Garis%20Panduan%20Audit%20Dokumentasi%20Klinikal%20dan%20Kod%20Klasifikasi.pdf [accessed 2026-08-27]
4. Asadi F, Ahmadi Hosseini M, Gomar T, Sabahi A. Factors affecting clinical coding errors. Shiraz E-Med J 2022;23(9):e122161. [doi: 10.5812/semj-122161]
5. Foster N, Raby E, Wood FM, Fear M, Pavlos N, Edgar DW. Evaluation of the accuracy of diagnostic coding and clinical documentation for traumatic heterotopic ossification diagnoses in Western Australian hospitals. Injury 2024 Mar;55(3):111329. [doi: 10.1016/j.injury.2024.111329] [Medline: 38296757]
6. Saizan S, Jaudin R, Yaacob NM, Sukeri S. The MalaysianDRG casemix system: financial implications of inaccurate clinical documentation and coding error. Malays J Med Health Sci 2021;17(1):83-87. [doi: 10.47836/mjmhs.v17i1.120]
7. Zafirah SA, Nur AM, Puteh SE, Aljunid SM. Potential loss of revenue due to errors in clinical coding during the implementation of the Malaysia diagnosis related group (MY-DRG®) Casemix system in a teaching hospital in Malaysia. BMC Health Serv Res 2018 Jan 25;18(1):38. [doi: 10.1186/s12913-018-2843-1] [Medline: 29370785]
8. Mahbubani K, Georgiades F, Goh EL, et al. Clinician-directed improvement in the accuracy of hospital clinical coding. Future Healthc J 2018 Feb;5(1):47-51. [doi: 10.7861/futurehosp.5-1-47] [Medline: 31098532]
9. Phooratsamee S, Treesak C, Mongkolsomlit S. Evaluating the accuracy and causes of discrepancies in recording disease codes and procedure codes, a case study of a community hospital in Lopburi Province. J Health Policy Law Adm 2025;11(3):685-695 [FREE Full text]
10. Agha L. The effects of health information technology on the costs and quality of medical care. J Health Econ 2014 Mar;34:19-30. [doi: 10.1016/j.jhealeco.2013.12.005] [Medline: 24463141]
11. Junttila JK, Koivu A, Fagerström L, Haatainen K, Nykänen P. Hospital mortality and optimality of nursing workload: a study on the predictive validity of the RAFAELA nursing intensity and staffing system. Int J Nurs Stud 2016 Aug;60:46-53. [doi: 10.1016/j.ijnurstu.2016.03.008] [Medline: 27297367]
12. Stone K, Zwiggelaar R, Jones P, Mac Parthaláin N. A systematic review of the prediction of hospital length of stay: towards a unified framework. PLOS Digit Health 2022 Apr;1(4):e0000017. [doi: 10.1371/journal.pdig.0000017] [Medline: 36812502]
13. Kuang A, Xu C, Southern DA, Sandhu N, Quan H. Validated administrative data based ICD-10 algorithms for chronic conditions: a systematic review. J Epidemiol Popul Health 2024 Aug;72(4):202744. [doi: 10.1016/j.jeph.2024.202744] [Medline: 38971056]



XSL•FO
RenderX

14. Upadhyay S, Hu HF. A qualitative analysis of the impact of electronic health records (EHR) on healthcare quality and safety: clinicians' lived experiences. Health Serv Insights 2022;15:11786329211070722. [doi: 10.1177/11786329211070722] [Medline: 35273449]
15. Bowman S. Impact of electronic health record systems on information integrity: quality and safety implications. Perspect Health Inf Manag 2013;10(Fall):1c. [Medline: 24159271]
16. Farzandipour M, Sheikhtaheri A, Sadoughi F. Effective factors on accuracy of principal diagnosis coding based on International Classification of Diseases, the 10th revision (ICD-10). Int J Inf Manag 2010 Feb;30(1):78-84. [doi: 10.1016/j.ijinfomgt.2009.07.002]
17. Maryati W, Murti B, Indarto D. Factors affecting the quality of diagnosis coding and medical record at Dr. Moewardi Hospital, Surakarta. J Health Policy Manag 2016;01(2):61-70. [doi: 10.26911/thejhpm.2016.01.02.01]
18. Firdayana F, Nurmalasari M, Hosizah H, Afrizal SH. Evaluation of the accuracy of medical terminology and its relationship to the accuracy of clinical coding in health facilities: systematic literature review. J Ners 2023;7(2):1171-1181. [doi: 10.31004/jn.v7i2.16381]
19. Alonso V, Santos JV, Pinto M, et al. Health records as the basis of clinical coding: is the quality adequate? A qualitative study of medical coders' perceptions. Health Inf Manag 2020 Jan;49(1):28-37. [doi: 10.1177/1833358319826351] [Medline: 30744403]
20. Pensions Act 1980. Laws of Malaysia. URL: https://world.moleg.go.kr/cms/commonDown.do?DLD_CFM_NO=PRWANH65CL3N4PNWUF71&FL_SEQ=151945 [accessed 2026-08-28]
21. Weidemann LA. Don't let DNFB cripple hospital cash flow. HealthLeaders Media. URL: https://www.healthleadersmedia.com/finance/dont-let-dnfb-cripple-hospital-cash-flow [accessed 2026-08-28]
22. Olagundoye O, van Boven K, Daramola O, Njoku K, Omosun A. Improving the accuracy of ICD-10 coding of morbidity/mortality data through the introduction of an electronic diagnostic terminology tool at the general hospitals in Lagos, Nigeria. BMJ Open Qual 2021 Mar;10(1):e000938. [doi: 10.1136/bmjoq-2020-000938] [Medline: 33674344]
23. Burns EM, Rigby E, Mamidanna R, et al. Systematic review of discharge coding accuracy. J Public Health (Oxf) 2012 Mar;34(1):138-148. [doi: 10.1093/pubmed/fdr054] [Medline: 21795302]
24. Taiwo Adeleke I, Ajayi OO, Jimoh AB, et al. Current clinical coding practices and implementation of ICD-10 in Africa: a survey of Nigerian hospitals. Am J Health Res 2015;3(1-1):38-46. [doi: 10.11648/j.ajhr.s.2015030101.16]
25. Nouraei SA, Virk JS, Hudovsky A, Wathen C, Darzi A, Parsons D. Accuracy of clinician-clinical coder information handover following acute medical admissions: implication for using administrative datasets in clinical outcomes management. J Public Health (Oxf) 2016 Jun;38(2):352-362. [doi: 10.1093/pubmed/fdv041] [Medline: 25907271]
26. Xu HA, Maccari B, Guillain H, Herzen J, Agri F, Raisaro JL. An end-to-end natural language processing application for prediction of medical case coding complexity: algorithm development and validation. JMIR Med Inform 2023 Jan 19;11:e38150. [doi: 10.2196/38150] [Medline: 36656627]
27. Chomutare T, Svenning TO, Hernández M, et al. Artificial intelligence to improve clinical coding practice in Scandinavia: crossover randomized controlled trial. J Med Internet Res 2025 Jul 3;27:e71904. [doi: 10.2196/71904] [Medline: 40608484]
28. Dong H, Falis M, Whiteley W, et al. Automated clinical coding: what, why, and where we are? NPJ Digit Med 2022 Oct 22;5(1):159. [doi: 10.1038/s41746-022-00705-7] [Medline: 36273236]
29. Campbell S, Giadresco K. Computer-assisted clinical coding: a narrative review of the literature on its benefits, limitations, implementation and impact on clinical coding professionals. Health Inf Manag 2020 Jan;49(1):5-18. [doi: 10.1177/1833358319851305] [Medline: 31159578]
30. Guidelines for achieving a compliant query practice (2022 update). AHIMA and ACDIS. 2022. URL: https://acdis.org/resources/guidelines-achieving-compliant-query-practice%E2%80%942022-update [accessed 2026-08-28]


## Abbreviations

**EMR:** electronic medical record
**HCTM:** Hospital Canselor Tuanku Muhriz
**ITCC:** International Centre for Casemix and Clinical Coding
**My-DRG:** Malaysian Diagnosis-Related Group

*Edited by M Balcarras; submitted 18.Feb.2026; peer-reviewed by MA Bujang, Y Chu; revised version received 26.Jun.2026; accepted 14.Jul.2026; published 17.Sep.2026.*

*Please cite as:*
*Azzeri A, Wan Musaludin WM, Jaafar H, Hafidz NMD, Noor Aizuddin A*
*Factors Affecting the Accuracy of Clinical Coding in the Casemix System in a Teaching Hospital: Cross-Sectional Study*
*Interact J Med Res 2026;15:e93769*
*URL: https://www.i-jmr.org/2026/1/e93769*
*doi:10.2196/93769*



XSL•FO
RenderX

© Amirah Azzeri, Wan Mastura Wan Musaludin, Hafiz Jaafar, Nurnabihah MD Hafidz, Azimatun Noor Aizuddin. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 17.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Individual Research Behaviors and Research Funding Acquisition Across Fields and Career Periods: Regression Analysis

Akiko Hashiguchi, PhD; Satoru Takahashi, MD, PhD

Planning Committee, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, Japan

**Corresponding Author:**
Akiko Hashiguchi, PhD
Planning Committee, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, Japan

## Abstract

**Background:** Acquiring external funding for academic research is a crucial element of researcher performance assessment. However, funding opportunities are often concentrated among a limited number of researchers, reinforcing existing structural inequalities. Under these conditions, it is important to clarify how individual researchers can improve their ability to secure external funding through their own actions and thereby build sustainable academic careers. In addition, fair faculty evaluation requires recognizing disciplinary differences in funding acquisition without assuming a normative research environment.

**Objective:** This study aimed to examine whether research-related behaviors at different career stages contribute to variations in funding acquisition capability across research fields.

**Methods:** We used funding data from Grant-in-Aid for Scientific Research, Japan's largest competitive research funding program, and classified 1152 professors according to their lifetime patterns of research funding acquisition. We also collected data on individual research-related behavioral characteristics, including supervisor prestige, authorship position, publication quality, collaborative research structure, and funding acquisition strategies, from publication records and administrative data, constructing a dataset of 804 researchers. By integrating these datasets, we conducted univariate and multivariate regression analyses to examine the factors associated with research funding acquisition.

**Results:** Field-specific differences were observed in the upper range of funding levels and individual behaviors. High-funding recipients were found only in the field of basic and internal medicine. Univariate logistic regression analyses showed that early-career publication success (basic medicine: odds ratio [OR] 1.48, 95% CI 1.22 - 1.79; internal medicine: OR 1.34, 95% CI 1.06 - 1.69; and surgery: OR 1.70, 95% CI 1.28 - 2.27) and midcareer senior authorship were significantly associated with funding acquisition (basic medicine: OR 5.99, 95% CI 2.63 - 13.65; internal medicine: OR 3.17, 95% CI 1.16 - 8.70; and surgery: OR 9.79, 95% CI 3.26 - 29.43). In addition, supporting specific collaborators (surgery: OR 2.09, 95% CI 1.10 - 3.97) and the quality of publications during the first 12 years of a researcher's career were major determinants of success (surgery: OR 2.22, 95% CI 1.18 - 4.17). In multiple linear regression analyses with total funding as the outcome, the adjusted $R^2$ values were 0.576 for surgery, 0.890 for social medicine, and 0.738 for dentistry, indicating that the models accounted for many behavioral factors associated with funding acquisition. However, the model fit suggests that additional relevant factors may influence funding acquisition in basic medicine, internal medicine, and nursing.

**Conclusions:** This study identified individual behaviors that influence Grant-in-Aid for Scientific Research funding acquisition across research fields. While organizational and academic structures have a significant impact, personal initiative during researchers' transition from peripheral to core roles becomes important, emphasizing strong relationship-building skills and outstanding publication records. As effective actions and their timing vary by field, individuals must consider their field characteristics and implement appropriate measures at suitable times.

***(Interact J Med Res 2026;15:e98428)*** doi:10.2196/98428

**KEYWORDS**

researcher performance; funding acquisition; Grant-in-Aid for Scientific Research (KAKENHI); research-related behaviors; collaborative research; research agenda reconfiguration

## Introduction

The increasing emphasis on quantifiable outcomes has normalized metric-driven evaluation in faculty career progression. Publishing productivity and acquisition of external funding are crucial elements in researcher performance assessment [1]. External funding not only enhances a university's reputation as an external evaluator of researchers' competence but also holds significant importance for university management under financial constraints [2].



XSL•FO
RenderX

The pronounced Matthew effect, whereby success in securing research funding leads to further success, should be taken into account when evaluating researchers' performance [3-5]. Since the 2000s, there has been a significant disparity in research funding acquisition, with a Gini coefficient exceeding 0.6, indicating that a disproportionate share of funding is concentrated among a small number of elite researchers [6]. Although total US health research funding has doubled, National Institutes of Health (NIH) Research Project Grants have become increasingly concentrated, with approximately 10% of funding going to a small number of elite researchers (the top 1%) [7]. A researcher's affiliated institution, past awards, and outstanding publication records create a "halo effect" that influences evaluators' judgments and, ultimately, determines the success or failure of research grant applications [8]. Previous studies have identified factors influencing funding acquisition, including the principal investigator's (PI) gender [9], age, academic rank [10], publication quality, international visibility [11], experience in collaborative research [12], and research interdisciplinarity [13]. These factors amplify and entrench initial, minor disparities throughout the course of a researcher's career. In other words, before researchers are able to lead original research grounded in their own experience, the success of their careers is strongly shaped by structural factors beyond their control.

Research on the "science of science" emphasizes that scientific success is structurally linked to embeddedness in elite academic networks rather than individual excellence [14,15]. The identified characteristics of outstanding researchers have led to policies aimed at fostering excellence, such as enhancing international mobility, promoting team science, and encouraging interdisciplinary collaboration. However, when considering the cumulative advantages, the factors regarded as determinants of scientific success can be understood as reflecting a researcher's position within organizational and academic structures—specifically, holding senior positions at prestigious universities. Consequently, research performance assessment based on currently available quantitative indicators may have limited value in assessing sustained and incremental academic contributions.

In parallel with growing recognition of persistent structural inequalities in academia, recent studies have increasingly examined the strategies through which researchers sustain their careers under such conditions. According to a study that analyzed the career development process at each stage in detail, researchers strategically exercised agency by adopting and constructing their own professional roles to secure the resources necessary to maximize their achievements as their careers progress [16]. In this context, investigating how the strategies adopted at different career stages contribute to the long-term sustainability of individual funding acquisition could provide insights for sustaining individual careers as well as maintaining the vitality of the research ecosystem [17].

Previous studies on academic success assumed a normative research environment and did not account for field-specific research practices. There is a well-known disparity between the humanities and natural sciences in publication and citation counts. In the medical field, the number of researchers exceeding a given threshold for the *h*-index varies by specialty [18], and the success rate of obtaining research grants as one's career progresses, particularly the transition from NIH-mentored career development (K) grants to independent investigator-initiated (R) grants, differs by specialty [19]. Disregard for field-specific contexts hinders fair evaluation within medical schools that encompass basic, clinical, and translational research. An accurate understanding of the differences in funding acquisition across fields is essential for valid researcher performance assessment.

Accordingly, this study investigated whether research-related behaviors at different career stages contribute to variations in funding acquisition capability across research fields.

## Methods

### Overview

In Japan, the Grant-in-Aid for Scientific Research (GIA) provides fundamental financial support for academic research in all fields. GIA offers a range of programs, including basic programs classified by the amount awarded, innovation programs that address new challenges, and specially promoted research programs that support outstanding original research and pioneer new areas of study. Funding allocation is based on evaluations of researchers' past achievements and the originality, scientific significance, feasibility, and soundness of proposed research projects. Therefore, a track record of securing GIA grants directly demonstrates a researcher's capabilities and future potential. For this reason, the amount of GIA funding received by each national university is a key indicator in university evaluations and is also a major metric for faculty recruitment and promotion.

### Data Source

Researchers in the medical field were extracted from the GIA database [20]. Given that GIA is effectively the only source of public funding scheme in Japan that supports investigator-initiated (bottom-up) research and that a track record of securing such grants is highly valued in recent faculty recruitment, it is reasonable to assume that this database covers nearly all researchers actively engaged in medical research at Japan's national universities, with the exception of those in the early stages of their careers or those whose primary duties are clinical work.

As many GIA-funded projects span 3 to 4 years, extracting researchers who initiated new projects in a given fiscal year enables us to identify approximately one-third of the population of scholars continuously engaged in GIA-funded research. Annual fluctuations in the amount and distribution of GIA grants reported by the funding agency are minimal; therefore, this sampling method yields a sample that is representative of the population.

In the GIA database, research institutions and implementation periods can be used as search terms to obtain information such as project titles, program categories, grant amounts, and researcher names. Information on the year of degree acquisition was obtained from the CiNii Research database [21] and used as an alternative indicator of researcher age, as exact ages were not publicly available. Each researcher's productivity and



XSL•FO
RenderX

publication quality were calculated from publication histories using data from the SciVal database, which is licensed by the University of Tsukuba, with the Scopus IDs as the search key [22].

## Ethical Considerations

This study did not require ethical approval because it did not involve data collection from human or animal subjects. Researchers' names and their grant statuses are publicly available. The GIA database, which is managed by the Ministry of Education, Culture, Sports, Science, and Technology in Japan, was created to ensure transparency in public research funding allocations and allow users to search for current research information in Japan. Researchers' publication statuses are available under the commercial license. The analysis plan was not registered because this study was a secondary analysis of data extracted from databases.

## Samples

We focused on medical school professors to understand how various activities at different career stages influence the success of research proposals over time. We identified 1526 professors as participants from national university medical schools who had secured new GIA projects in 2024 (Table 1). To understand the career stages of these professors at the time they received GIA grants, we analyzed the years in which they obtained their doctoral degrees. This allowed us to identify the doctoral graduation years and institutions of 1157 professors (Table 1). We then used the number of years since degree completion to examine when and through which programs these professors received project grants, thereby identifying patterns in GIA acquisition across career stages.

**Table .** Researcher population and their characteristics.

| Population | N | Female, n (%) | Field, n (%) | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Basic medicine | Clinical medicine (internal medicine) | Clinical medicine (surgery) | Social medicine | Nursing | Dentistry |
| Initial sample ($N_0$) | 1526 | 187 (12.25) | 424 (27.79) | 283 (18.48) | 446 (29.23) | 105 (6.88) | 152 (9.96) | 116 (7.60) |
| Degree matched ($N_1$) | 1152 | 87 (7.55) | 369 (32.03) | 229 (19.88) | 338 (29.34) | 62 (5.38) | 59 (5.12) | 95 (8.25) |
| Publication record retrieved ($N_2$) | 804 | 68 (8.46) | 240 (29.85) | 162 (20.14) | 247 (30.72) | 48 (5.97) | 54 (6.71) | 53 (6.59) |

We further linked the data for the 1157 professors to behavioral indicators derived from their publication records. As SciVal provides publication data only from 1996 onward, the analysis was restricted to researchers who obtained their doctoral degrees in or after 1996 to capture publication activities from the early stages of their careers. Using SciVal, we successfully identified publication histories for 804 professors. For the calculation of behavioral indicators, we included all publication types rather than restricting the analysis to research articles alone. This approach was intended to capture diverse forms of scholarly contribution and account for the increasing diversity of publication practices accompanying the expansion of interdisciplinary research. Subsequent analyses were therefore conducted using this subset (Table 1).

## Metrics

The metrics used in this study are as follows (Multimedia Appendix 1). For researcher competitiveness, we calculated the total grant amount secured as a PI (total funding) and the total grant amount secured as a PI during the 12 years following the completion of a doctoral degree (funding_0–12y). The latter was included as a lagged variable to mitigate endogeneity.

Individual attributes include the number of years since obtaining a doctoral degree (career length), gender (gender), and the rank of the university where each researcher obtained their doctoral degree (university rank [graduated]) and the university with which they are currently affiliated (university rank [currently affiliated]).

To evaluate the quality of research output and leadership throughout a researcher's career, we calculated several metrics based on publication data: the total number of publications (total publication counts); the proportion of publications in which the researcher was listed as the first, second, or senior author (% lead-authored publications); and the average Source Normalized Impact per Paper (SNIP) of lead-authored publications (average SNIP). Journal SNIP scores were used as a proxy for research quality rather than paper-level metrics because they demonstrate that a paper meets the journal's publication standards. These metrics indicate the researcher's overall scholarly strength.

Furthermore, to clarify research activities and achievements during the initial career stage, we quantified the mentor's competitiveness (mentor level). On the basis of the organizational structure of GIA projects, we defined a participant's mentor as the PI who first appointed the participant as a co-investigator in a GIA project. Mentor level was quantified using the largest GIA awarded to that mentor. Additionally, we calculated the number of first-authored publications produced within the first 5 years after degree completion (#first-authored_0–5y) and the total SNIP of those publications (SNIP_0–5y).



XSL•FO
RenderX

To identify participants' strengths as collaborators and leaders within research networks, we calculated the following two metrics: (1) the ratio of total to unique PIs across projects involving the participant as a project member, reflecting the extent to which a researcher is repeatedly recruited by the same PI, capturing the researcher's value and reliability as a research collaborator and their access to established research environments (elite reliance); and (2) the ratio of total to unique collaborators across projects secured by the participant as a PI, reflecting the extent to which a researcher repeatedly recruits the same collaborators across their own projects, capturing the ability to build and sustain collaborative research networks (selective mobilization ability). These 2 indicators were conceived based on the findings suggesting that early-career researchers enhance their competitiveness by gaining experience in large-scale projects led by established researchers, while established researchers improve their research outcomes by connecting projects involving diverse participants under a broad conceptual framework [23,24].

Regarding researchers' funding acquisition strategies, we calculated the number of research projects secured as a PI (# projects); the number of research projects secured as a PI in large-scale research programs (Grant-in-Aid for Specially Promoted Research, Grants-in-Aid for Scientific Research (S), and (A)) (# large-scale projects); and the proportion of research grants acquired as a PI in the category of Challenging Exploratory Research (which, following institutional reforms, includes Grant-in-Aid for Exploratory Research, Challenging Exploratory Research, and Challenging Research [Pioneering] or [Exploratory]) (% challenging research). In this context, engagement in large-scale research programs was interpreted as indicating commitment to a core research agenda, whereas participation in challenging exploratory research was interpreted as indicating engagement with complementary research agendas.

Finally, to examine research performance across different career stages, we divided academic careers into 3 periods according to years since degree completion: early career (0 - 12 y), midcareer (13 - 24 y), and late career (≥25 y). For each stage, we calculated the average SNIP of publications in which the researcher served as the first, second, or senior author. Indicator names were defined according to author position and career stage. For example, SNIP_first_0–12y denotes the average SNIP of first-authored publications during the early career stage, whereas SNIP_second_13–24y and SNIP_senior_25+y denote the corresponding indicators for second-authored publications during the midcareer stage and senior-authored publications during the late career stage, respectively.

### Statistical Analysis

Agglomerative hierarchical clustering was performed using XLSTAT statistical software (Addinsoft) to group the participants. Univariate logistic regression and multiple linear regression analyses were conducted using XLSTAT to estimate the impact of factors such as career length, university ranking, professional networks, productivity, output quality, leadership in academic writing, and grant acquisition strategies on GIA acquisition.

## Results

### Researcher Types Based on Funding Acquisition Status and Distribution by Field

Table 2 presents the descriptive statistics of the variables used in the analysis. For agglomerative hierarchical clustering, we calculated the total GIA funding acquired throughout each researcher's career as a PI and funding acquired during the first 12 years. Researchers (N=1152) were grouped based on total funding (indicator: total funding), early career funding (indicator: funding_0–12y), and career duration (indicator: career length). Six phenotype clusters were identified (Table 3). On the basis of the total funding and acquisition patterns (Multimedia Appendix 2), the groups were classified as follows: group 1, established type; group 2, emerging type; group 3, catching-up trajectory type; group 4, plateaued type; group 5, moderate-level stable type; and group 6, low-slope stable type.



XSL•FO
RenderX

**Table .** Descriptive statistics.

| Metrics | Degree matched ($N_1$=1152) | | | | Publication record retrieved ($N_2$=804) | | | |
|---|---|---|---|---|---|---|---|---|
| | Minimum | Maximum | Median (IQR) | Mean (SD) | Minimum | Maximum | Median | Mean (SD) |
| Total funding (million US $[a]) | 0.01 | 16.02 | 0.35 (0.17 - 0.90) | 0.99 (1.93) | 0.00 | 7.17 | 0.27 (0.16 - 0.77) | 0.83 (0.51) |
| Funding_0–12y (million US $[a]) | 0.00 | 7.17 | 0.10 (0.03 - 0.20) | 0.18 (0.44) | 0.01 | 15.44 | 0.10 (0.04 - 0.20) | 0.21 (1.68) |
| Career length (y) | 11.00 | 48.00 | 26.00 (21.00 - 30.00) | 25.55 (5.99) | 11.00 | 29.00 | 23.00 (20.00 - 26.00) | 22.76 (4.33) |
| University rank (currently affiliated)[b] | 0.00 | 1.00 | 0.00 (0.00 - 0.00) | 0.08 (0.26) | 1.00 | 7.00 | 5.00 (2.00 - 5.00) | 4.20 (1.74) |
| University rank (graduated)[b] | 1.00 | 7.00 | 4.00 (2.00 - 5.00) | 4.18 (1.76) | 1.00 | 7.00 | 3.00 (2.00 - 5.00) | 3.25 (1.83) |
| # projects | — [c] | — | — | — | 1.00 | 37.00 | 6.00 (4.00 - 8.00) | 6.64 (4.06) |
| # large-scale projects | — | — | — | — | 0.00 | 6.00 | 0.00 (0.00 - 0.00) | 0.16 (0.58) |
| % challenging research | — | — | — | — | 0.00 | 4.00 | 0.00 (0.00 - 0.25) | 0.16 (0.33) |
| Total publication counts | — | — | — | — | 1.00 | 901.00 | 123.50 (72.00 - 207.00) | 159.46 (129.85) |
| % lead-authored publications | — | — | — | — | 0.00 | 1.64 | 0.53 (0.40 - 0.68) | 0.55 (0.22) |
| Average SNIP[d] | — | — | — | — | 0.00 | 3.26 | 0.78 (0.62 - 0.99) | 0.83 (0.36) |
| # first-authored_0–5y | — | — | — | — | 0.00 | 49.00 | 3.00 (1.00 - 5.00) | 3.56 (4.14) |
| SNIP_0–5y | — | — | — | — | 0.00 | 12.18 | 1.37 (0.51 - 2.05) | 1.54 (1.50) |
| SNIP_first_0–12y | — | — | — | — | 0.00 | 13.22 | 0.67 (0.37 - 1.07) | 0.84 (0.85) |
| SNIP_first_13–24y | — | — | — | — | 0.00 | 3.83 | 0.00 (0.00 - 0.07) | 0.12 (0.32) |
| SNIP_first_25+y | — | — | — | — | 0.00 | 14.59 | 0.68 (0.23 - 1.53) | 1.29 (1.86) |
| SNIP_second_0–12y | — | — | — | — | 0.00 | 7.21 | 1.01 (0.76 - 1.25) | 1.05 (0.62) |
| SNIP_second_13–24y | — | — | — | — | 0.00 | 10.61 | 0.00 (0.00 - 0.70) | 0.36 (0.69) |
| SNIP_second_25+y | — | — | — | — | 0.00 | 8.43 | 1.00 (0.68 - 1.28) | 1.02 (0.74) |
| SNIP_senior_0–12y | — | — | — | — | 0.00 | 10.26 | 0.81 (0.33 - 1.09) | 0.78 (0.67) |
| SNIP_senior_13–24y | — | — | — | — | 0.00 | 1.59 | 0.00 (0.00 - 0.00) | 0.08 (0.21) |



XSL•FO
RenderX

| Metrics | Degree matched ($N_1$=1152) | | | | Publication record retrieved ($N_2$=804) | | | |
|---|---|---|---|---|---|---|---|---|
| | Minimum | Maximum | Median (IQR) | Mean (SD) | Minimum | Maximum | Median | Mean (SD) |
| SNIP_senior_25+y | — | — | — | — | 0.00 | 3.86 | 0.35 (0.00 - 1.00) | 0.56 (0.66) |
| Mentor level[e] | — | — | — | — | 1.00 | 5.00 | 5.00 (5.00 - 5.00) | 4.71 (0.86) |
| Elite reliance[f] | — | — | — | — | 1.00 | 10.00 | 1.37 (1.07 - 1.67) | 1.51 (0.68) |
| Selective mobilization capacity[g] | — | — | — | — | 0.00 | 6.00 | 1.25 (1.00 - 1.60) | 1.30 (0.63) |

[a]A currency exchange rate of JP ¥1=US $0.0067 is applicable.

[b]These variables were quantified based on The Times Higher Education World University Rankings 2023.

[c]Not applicable.

[d]SNIP: Source Normalized Impact per Paper.

[e]These variables were quantified based on the size of the Grant-in-Aid for Scientific Research projects secured by the mentor, with smaller values indicating greater competitiveness. The categories are coded as 1=Grant-in-Aid for Scientific Research on Innovative Areas, 2=the Grant-in-Aid for Specially Promoted Research, 3=Grant-in-Aid for Scientific Research (S), 4=Grant-in-Aid for Scientific Research (A), and 5=Grant-in-Aid for Scientific Research (B) or smaller.

[f]The ratio of the total number of principal investigator (PI)–participant collaborations to the number of distinct PIs across all projects in which the participant served as a project member. Larger values indicate higher level of reliance.

[g]The ratio of the total number of collaborator participations to the number of distinct collaborators across projects secured by the participant as a PI. The closer the value is to 1, the greater the team's diversity; conversely, higher values indicate stronger ability to mobilize specific collaborators.

**Table .** Grant acquisition status as a principal investigator by phenotype clusters.

| Metrics | All | Phenotype clusters | | | | | |
|---|---|---|---|---|---|---|---|
| | | Group 1 | Group 2 | Group 3 | Group 4 | Group 5 | Group 6 |
| $N_1$ (n [female], %) | 1152 (87, 7.55) | 11 (1, 9.1) | 3 (0, 0.0) | 26 (1, 3.8) | 46 (3, 6.5) | 295 (21, 7.12) | 771 (61, 7.91) |
| Career length (y), median (IQR) | 26 (21 - 30) | 32 (29 - 31) | 22 (17 - 20) | 33 (27 - 30) | 31 (22 - 27) | 32 (23 - 27) | 29 (21 - 25) |
| Funding_0–12y (million US $[a]), median (IQR) | 0.09 (0.03 - 0.19) | 0.00 (0.00 - 0.00) | 6.65 (5.32 - 6.69) | 0.00 (0.00 - 0.31) | 0.00 (0.00 - 0.42) | 0.25 (0.00 - 0.47) | 0.06 (0.03 - 0.13) |
| Total funding (million US $[a]), median (IQR) | 0.33 (0.16 - 0.85) | 11.62 (11.00 - 14.10) | 9.80 (8.74 - 10.00) | 7.45 (6.64 - 8.28) | 4.72 (3.83 - 4.82) | 1.00 (0.82 - 1.26) | 0.20 (0.13 - 0.33) |

[a]A currency exchange rate of JP ¥1=US $0.0067 is applicable.

The distribution of personal attributes and research fields by phenotype cluster is summarized in Table 4. No differences were observed between clusters in gender, but a significant difference was found in the distribution of university rank (Fisher exact test, $P$<.05). Notably, participants in groups 1, 2, and 3, characterized by high funding acquisition capabilities, appeared regardless of the university ranks (graduated); within group 1, a total of 4 participants had graduated from top-tier universities, 3 had graduated from midtier universities, and 4 had graduated from lower-tier universities (Table 4).



XSL•FO
RenderX

**Table .** Personal attributes and research fields by phenotype cluster.

| Metrics | All ($N_1$=1152), n (%) | Phenotype clusters, n (%) | | | | | | *P* value (Fisher exact test[b]) |
|---|---|---|---|---|---|---|---|---|
| | | Group 1 (n=11)[a] | Group 2 (n=3)[a] | Group 3 (n=26)[a] | Group 4 (n=46)[a] | Group 5 (n=295)[a] | Group 6 (n=771)[a] | |
| Gender | | | | | | | | .96 |
| Male | 1065 (92.45) | 10 (91.0) | 3 (100.0) | 25 (96.2) | 43 (93.5) | 274 (92.88) | 710 (92.09) | |
| Female | 87 (7.55) | 1 (9.0) | 0 (0) | 1 (3.8) | 3 (6.5) | 21 (7.12) | 61 (7.91) | |
| University rank (currently affiliated)[c] | | | | | | | | <.001 |
| High (≤100) | 51 (4.43) | 4 (36.4) | 0 (0) | 4 (15.4) | 5 (10.9) | 22 (7.46) | 16 (2.08) | |
| Medium (101 - 600) | 256 (22.22) | 5 (45.5) | 2 (66.7) | 10 (38.5) | 17 (37.0) | 80 (27.12) | 142 (18.42) | |
| Low (≥600) | 845 (73.35) | 2 (18.2) | 1 (33.3) | 12 (46.2) | 24 (52.2) | 193 (65.42) | 613 (79.51) | |
| University rank (graduated)[c] | | | | | | | | <.001 |
| High (≤100) | 178 (15.45) | 4 (36.4) | 1 (33.3) | 9 (34.6) | 11 (23.9) | 72 (24.41) | 81 (10.51) | |
| Medium (101 - 600) | 383 (33.25) | 3 (27.3) | 1 (33.3) | 11 (42.3) | 19 (41.3) | 120 (40.68) | 229 (29.7) | |
| Low (≥600) | 591 (51.3) | 4 (36.4) | 1 (33.3) | 6 (23.1) | 16 (34.8) | 103 (34.92) | 461 (59.79) | |
| Fields | | | | | | | | <.001 |
| Basic medicine | 369 (32.03) | 9 (81.8) | 1 (33.3) | 19 (73.1) | 32 (69.6) | 115 (38.98) | 193 (25.03) | |
| Clinical medicine (internal medicine) | 229 (19.88) | 2 (18.2) | 1 (33.3) | 3 (11.5) | 8 (17.4) | 42 (14.24) | 173 (22.44) | |
| Clinical medicine (surgery) | 338 (29.34) | 0 (0) | 0 (0) | 3 (11.5) | 3 (6.5) | 80 (27.12) | 252 (32.68) | |
| Social medicine | 62 (5.38) | 0 (0) | 1 (33.3) | 0 (0) | 1 (2.2) | 15 (5.08) | 45 (5.84) | |
| Nursing | 59 (5.12) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 9 (3.05) | 50 (6.49) | |
| Dentistry | 95 (8.25) | 0 (0) | 0 (0) | 1 (3.8) | 2 (4.3) | 34 (11.53) | 58 (7.52) | |

[a]The characteristics and median total funding amounts for each group are as follows: group 1, established type, US $11.62 million; group 2, emerging type, US $9.80 million; group 3, catching-up trajectory type, US $7.45 million; group 4, plateaued type, US $4.72 million; group 5, moderate-level stable type, US $1.00 million; and group 6, low-slope stable type, US $0.20 million.

[b]Groups 1 and 2 were combined because of sparse cell counts. *P* values were calculated using Fisher exact test with Monte Carlo simulation (5000 simulations).

[c]These variables were quantified based on The Times Higher Education World University Rankings 2023.

Research fields were categorized into 6 areas: basic medicine, clinical medicine (internal medicine), clinical medicine (surgery), social medicine, nursing, and dentistry (Table 4). Basic medicine and clinical medicine (both internal medicine and surgery) accounted for the largest proportion of participants because the analysis focused on professors affiliated with medical schools and excluded researchers from independent nursing or dental schools. Public health studies are less developed in Japan than in some other countries and do not constitute an independent faculty member. The distribution of researchers securing high funding varied by field (Fisher exact test, $P<.01$). Group 1 participants were only observed in basic medicine and internal medicine (Table 4). In surgery and dentistry, group 3 received the highest total funding, whereas in social medicine, group 2 ranked highest (Table 4). None of the nursing researchers secured high levels of research funding, defined as group 4 or higher, with a median total grant amount of US $4.72 (IQR US $3.83 - US $4.82) million (JP ¥1 [US $0.0067]). These results suggest field-specific differences in the upper funding level ranges.

To examine differences in behavioral characteristics by field, we compiled indicator values (n=804; Table 5). Significant differences were found across fields for all indicators, revealing distinct research practices between fields (ANOVA, $P<.05$). Basic medicine and internal medicine were characterized by high-quality outcomes early in careers (SNIP_0–5y: basic



XSL•FO
RenderX

medicine, mean 1.86, SE 0.11; and internal medicine: mean 1.86, SE 0.12) and throughout their careers (average SNIP: basic medicine: mean 0.94, SE 0.03; internal medicine: mean 0.89, SE 0.03), while basic medicine secured more large-scale research grants (# large-scale projects: mean 0.28, SE 0.05). Internal medicine, surgery, and social medicine demonstrated high productivity (total publication counts: internal medicine, mean 197.58, SE 11.11; surgery, mean 189.33, SE 8.88; social medicine, mean 199.40, SE 22.70). Social medicine was characterized by high initial productivity (# first-authored_0–5y: mean 5.19, SE 1.12), high collaborator diversity—indicated by low values for the following metrics (elite reliance: mean 1.34, SE 0.05; selective mobilization capacity: mean 1.15, SE 0.08)—and frequent implementation of large-scale projects (# large-scale projects: mean 0.25, SE 0.09). The use of complementary research programs was more common among nurses (% challenging research: mean 0.26, SE 0.08). In dentistry, the same researchers frequently collaborated on multiple projects (as indicated by high values for elite reliance: mean 1.72, SE 0.09 and selective mobilization capacity: mean 1.49, SE 0.09).

**Table .** Behavioral characteristics by fields ($N_2$=804).

| Metrics | Phenotype clusters, mean (SE) | | | | | | ANOVA *P* value |
|---|---|---|---|---|---|---|---|
| | Basic medicine (n=240) | Clinical medicine (internal medicine; n=162) | Clinical medicine (surgery; n=247) | Social medicine (n=48) | Nursing (n=54) | Dentistry (n=53) | |
| Mentor level[a] | 4.43 (0.07) | 4.7 (0.07) | 4.87 (0.04) | 4.88 (0.09) | 4.94 (0.03) | 4.77 (0.1) | <.001 |
| # first-authored_0–5y | 3.6 (0.21) | 3.96 (0.37) | 3.67 (0.25) | 5.19 (1.12) | 1.3 (0.28) | 2.45 (0.35) | <.001 |
| SNIP_0–5y | 1.86 (0.11) | 1.86 (0.12) | 1.35 (0.08) | 1.41 (0.2) | 0.44 (0.1) | 1.22 (0.16) | <.001 |
| Elite reliance | 1.62 (0.06) | 1.48 (0.04) | 1.45 (0.03) | 1.34 (0.05) | 1.38 (0.05) | 1.72 (0.09) | .002 |
| Selective mobilization capacity | 1.21 (0.04) | 1.28 (0.06) | 1.38 (0.03) | 1.15 (0.08) | 1.38 (0.07) | 1.49 (0.09) | .004 |
| Total publication counts | 128.58 (6.34) | 197.97 (11.11) | 189.33 (8.88) | 199.40 (22.70) | 48.46 (9.19) | 119.3 (10.48) | <.001 |
| % lead-authored publications | 0.56 (0.01) | 0.52 (0.01) | 0.53 (0.01) | 0.51 (0.03) | 0.69 (0.05) | 0.55 (0.03) | <.001 |
| Average SNIP[b] | 0.94 (0.03) | 0.89 (0.03) | 0.75 (0.02) | 0.79 (0.04) | 0.62 (0.05) | 0.82 (0.04) | <.001 |
| # projects | 8.20 (0.31) | 6.49 (0.33) | 5.68 (0.19) | 5.25 (0.46) | 5.07 (0.34) | 7.32 (0.5) | <.001 |
| # large-scale projects | 0.28 (0.05) | 0.12 (0.04) | 0.09 (0.03) | 0.25 (0.09) | 0.11 (0.05) | 0 (0) | .001 |
| % challenging research | 0.12 (0.02) | 0.14 (0.02) | 0.19 (0.02) | 0.18 (0.04) | 0.26 (0.07) | 0.16 (0.04) | .04 |

[a]These variables were quantified based on the size of the Grant-in-Aid for Scientific Research projects secured by the mentor, with smaller values indicating greater competitiveness. The categories are coded as 1=Grant-in-Aid for Scientific Research on Innovative Areas, 2=the Grant-in-Aid for Specially Promoted Research, 3=Grant-in-Aid for Scientific Research (S), 4=Grant-in-Aid for Scientific Research (A), and 5=Grant-in-Aid for Scientific Research (B) or smaller.

[b]SNIP: Source Normalized Impact per Paper.

## Associations Between Behavioral Characteristics and Funding Acquisition Capabilities by Field

We examine the association between behavioral characteristics and funding acquisition using univariate logistic regression analyses conducted separately for each field (Table 6). Researchers in group 6 were used as the reference group, and all other types were pooled into a single comparison group (Table 6). Several common patterns emerged regarding the factors associated with substantial funding acquisition.



XSL•FO
RenderX

**Table .** Associations between behavioral characteristics and funding acquisition by fields (2-tailed $\chi^2$ test)[a].

| Variable and fields | Odds ratio (95% CI) | Significance of the model, *P* value (Wald test) |
|---|---|---|
| Mentor level ($N_2$=804)[b] | | |
| Basic medicine (n=240) | 0.49 (0.36 - 0.66)[c] | <.001[c] |
| Internal medicine (n=162) | 0.67 (0.46 - 0.96)[d] | .03[d] |
| Surgery (n=247) | 0.38 (0.22 - 0.68)[c] | .001[e] |
| Social medicine (n=48) | 0.25 (0.02 - 2.70) | .26 |
| Nursing (n=54) | 0.32 (0.03 - 3.99) | .38 |
| Dentistry (n=53) | 0.63 (0.29 - 1.39) | .25 |
| # first-authored_0–5y ($N_2$=804) | | |
| Basic medicine (n=240) | 1.06 (0.98 - 1.15) | .14 |
| Internal medicine (n=162) | 1.02 (0.95 - 1.10) | .58 |
| Surgery (n=247) | 1.00 (0.92 - 1.08) | .98 |
| Social medicine (n=48) | 1.02 (0.94 - 1.10) | .65 |
| Nursing (n=54) | 1.18 (0.86 - 1.61) | .31 |
| Dentistry (n=53) | 0.96 (0.75 - 1.24) | .78 |
| SNIP_0–5y[f] ($N_2$=804) | | |
| Basic medicine (n=240) | 1.48 (1.22 - 1.79)[c] | <.001[c] |
| Internal medicine (n=162) | 1.34 (1.06 - 1.69)[d] | .02[d] |
| Surgery (n=247) | 1.70 (1.28 - 2.27)[c] | .000[c] |
| Social medicine (n=48) | 1.6 (0.98 - 2.59) | .06 |
| Nursing (n=54) | 1.7 (0.65 - 4.44) | .28 |
| Dentistry (n=53) | 1.58 (0.88 - 2.83) | .13 |
| Elite reliance ($N_2$=804) | | |
| Basic medicine (n=240) | 0.99 (0.76 - 1.28) | .92 |
| Internal medicine (n=162) | 1.20 (0.63 - 2.29) | .58 |
| Surgery (n=247) | 2.09 (1.10 - 3.97)[d] | .03[d] |
| Social medicine (n=48) | 9.09 (1.32 - 62.57)[d] | .03[d] |
| Nursing (n=54) | 6.60 (1.02 - 42.66)[d] | .047[d] |
| Dentistry (n=53) | 1.01 (0.4 - 2.54) | .99 |
| Selective mobilization capacity ($N_2$=804) | | |
| Basic medicine (n=240) | 1.15 (0.78 - 1.71) | .48 |
| Internal medicine (n=162) | 1.44 (0.91 - 2.28) | .12 |
| Surgery (n=247) | 1.84 (1.00 - 3.38) | .05 |
| Social medicine (n=48) | 2.91 (0.71 - 11.89) | .14 |
| Nursing (n=54) | 2.77 (0.58 - 13.29) | .20 |
| Dentistry (n=53) | 5.98 (1.42 - 25.21)[d] | .01[d] |
| Total publication count ($N_2$=804) | | |
| Basic medicine (n=240) | 1.00 (1.00 - 1.00) | .68 |
| Internal medicine (n=162) | 1.00 (1.00 - 1.00) | .26 |
| Surgery (n=247) | 1.00 (0.998 - 1.003) | .70 |



XSL•FO
RenderX

| Variable and fields | Odds ratio (95% CI) | Significance of the model, *P* value (Wald test) |
|---|---|---|
| Social medicine (n=48) | 1.01 (1.00 - 1.01)[d] | .03[d] |
| Nursing (n=54) | 1.01 (1.00 - 1.02)[d] | .04[d] |
| Dentistry (n=53) | 1.00 (1.00 - 1.01) | .57 |
| % lead-authored publications ($N_2$=804) | | |
| Basic medicine (n=240) | 2.07 (0.64 - 6.69) | .23 |
| Internal medicine (n=162) | 4.47 (0.61 - 32.78) | .14 |
| Surgery (n=247) | 2.5 (0.58 - 10.79) | .22 |
| Social medicine (n=48) | 3.11 (0.14 - 67.89) | .47 |
| Nursing (n=54) | 1.53 (0.21 - 11.28) | .67 |
| Dentistry (n=53) | 5.09 (0.17 - 152.02) | .35 |
| Average SNIP ($N_2$=804) | | |
| Basic medicine (n=240) | 5.99 (2.63 - 13.65)[c] | <.001[c] |
| Internal medicine (n=162) | 3.17 (1.16 - 8.70)[d] | .03[d] |
| Surgery (n=247) | 9.79 (3.26 - 29.43)[c] | <.001[c] |
| Social medicine (n=48) | 1.00 (0.09 - 11.84) | .998 |
| Nursing (n=54) | 0.71 (0.07 - 6.98) | .77 |
| Dentistry (n=53) | 10.76 (1.09 - 105.96)[d] | .04[d] |
| # large-scale projects ($N_2$=804) | | |
| Basic medicine (n=240) | 25.41 (3.53 - 183.03)[c] | .001[c] |
| Internal medicine (n=162) | __[g] | .99 |
| Surgery (n=247) | 24.27 (5.16 - 114.09)[c] | <.001[c] |
| Social medicine (n=48) | 9.67 (1. 98 - 47.34)[c] | .005[e] |
| Nursing (n=54) | __[g] | .99 |
| Dentistry (n=53) | __[g] | .99 |
| % challenging research ($N_2$=804) | | |
| Basic medicine (n=240) | 0.66 (0.26 - 1.71) | .40 |
| Internal medicine (n=162) | 0.59 (0.10 - 3.56) | .57 |
| Surgery (n=247) | 2.03 (0.90 - 4.60) | .09 |
| Social medicine (n=48) | 12.67 (1.13 - 142.10)[d] | .04[d] |
| Nursing (n=54) | 0.68 (0.11 - 4.25) | .68 |
| Dentistry (n=53) | 1.37 (0.23 - 8.30) | .73 |
| SNIP_first_0 - 12y ($N_2$=804) | | |
| Basic medicine (n=240) | 1.17 (0.88 - 1.57) | .28 |
| Internal medicine (n=162) | 1.00 (0.62 - 1.63) | .99 |
| Surgery (n=247) | 1.56 (1.04 - 2.35)[d] | .03[d] |
| Social medicine (n=48) | 0.97 (0.29 - 3.27) | .97 |
| Nursing (n=54) | 1.12 (0.46 - 2.72) | .81 |
| Dentistry (n=53) | 1.10 (0.61 - 1.99) | .76 |
| SNIP_first_13 - 24y ($N_2$=804) | | |



XSL•FO
RenderX

| Variable and fields | Odds ratio (95% CI) | Significance of the model, *P* value (Wald test) |
|---|---|---|
| Basic medicine (n=240) | 0.89 (0.45 - 1.77) | .74 |
| Internal medicine (n=162) | 1.10 (0.3 - 4.10) | .88 |
| Surgery (n=247) | 3.01 (0.89 - 10.14) | .08 |
| Social medicine (n=48) | 4.82 (0.36 - 64.39) | .23 |
| Nursing (n=54) | 6.87 (0.62 - 76.69) | .12 |
| Dentistry (n=53) | 0.60 (0.13 - 2.84) | .52 |
| SNIP_first_25+y ($N_2$=804) | | |
| Basic medicine (n=240) | 1.11 (0.98 - 1.25) | .10 |
| Internal medicine (n=162) | 1.10 (0.92 - 1.31) | .31 |
| Surgery (n=247) | 0.96 (0.76 - 1.20) | .70 |
| Social medicine (n=48) | 0.92 (0.61 - 1.39) | .68 |
| Nursing (n=54) | 0.85 (0.22 - 3.33) | .81 |
| Dentistry (n=53) | 1.00 (0.71 - 1.41) | .99 |
| SNIP_second_0 - 12 y ($N_2$=804) | | |
| Basic medicine (n=240) | 1.79 (1.16 - 2.75) | .008 |
| Internal medicine (n=162) | 3.31 (1.6 - 6.86)[c] | .001[c] |
| Surgery (n=247) | 2.22 (1.18 - 4.17)[d] | .01[d] |
| Social medicine (n=48) | 2.50 (0.56 - 11.11) | .23 |
| Nursing (n=54) | 1.55 (0.38 - 6.30) | .54 |
| Dentistry (n=53) | 2.2 (0.68 - 7.16) | .19 |
| SNIP_second_13 - 24 y ($N_2$=804) | | |
| Basic medicine (n=240) | 1.36 (0.85 - 2.17) | .21 |
| Internal medicine (n=162) | 0.77 (0.44 - 1.33) | .35 |
| Surgery (n=247) | 1.11 (0.67 - 1.81) | .69 |
| Social medicine (n=48) | 1.61 (0.52 - 5.03) | .41 |
| Nursing (n=54) | 1.45 (0.27 - 7.70) | .67 |
| Dentistry (n=53) | 1.42 (0.40 - 5.05) | .59 |
| SNIP_second_25+ y ($N_2$=804) | | |
| Basic medicine (n=240) | 2.43 (1.59 - 3.70)[c] | <.001[c] |
| Internal medicine (n=162) | 1.09 (0.75 - 1.58) | .66 |
| Surgery (n=247) | 2.17 (1.16 - 4.04)[d] | .02[d] |
| Social medicine (n=48) | 1.60 (0.44 - 5.80) | .47 |
| Nursing (n=54) | 1.60 (0.42 - 6.01) | .49 |
| Dentistry (n=53) | 4.80 (1.07 - 21.55)[d] | .04[d] |
| SNIP_senior_0 - 12y ($N_2$=804) | | |
| Basic medicine (n=240) | 3.12 (1.89 - 5.16)[c] | <.001[c] |
| Internal medicine (n=162) | 2.61 (1.43 - 4.76)[e] | .002[e] |
| Surgery (n=247) | 2.03 (1.15 - 3.58)[d] | .02[d] |
| Social medicine (n=48) | 11.66 (1.75 - 77.73)[d] | .01[d] |
| Nursing (n=54) | 1.96 (0.55 - 6.95) | .30 |



XSL•FO
RenderX

| Variable and fields | Odds ratio (95% CI) | Significance of the model, *P* value (Wald test) |
|---|---|---|
| Dentistry (n=53) | 1.75 (0.58 - 5.29) | .32 |
| SNIP_senior_13 - 24y ($N_2$=804) | | |
| Basic medicine (n=240) | 2.49 (0.71 - 8.81) | .16 |
| Internal medicine (n=162) | 1.20 (0.27 - 5.39) | .81 |
| Surgery (n=247) | 1.39 (0.35 - 5.55) | .64 |
| Social medicine (n=48) | 7.28 (0.35 - 151.28.20) | .20 |
| Nursing (n=54) | 2.12 (0.01 - 314.94) | .77 |
| Dentistry (n=53) | 1.40 (0.14 - 13.59) | .77 |
| SNIP_senior_25+y ($N_2$=804) | | |
| Basic medicine (n=240) | 2.15 (1.46 - 3.17)[c] | <.001[c] |
| Internal medicine (n=162) | 2.09 (1.27 - 3.45)[e] | .004[e] |
| Surgery (n=247) | 1.77 (1.06 - 2.96)[d] | .03[d] |
| Social medicine (n=48) | 1.15 (0.39 - 3.39) | .80 |
| Nursing (n=54) | 0.74 (0.11 - 5.18) | .76 |
| Dentistry (n=53) | 0.82 (0.29 - 2.3) | .70 |

[a]The researcher type was used as the dependent variable. Group 6 was used as a reference.

[b]These variables were quantified based on the largest grant category secured by the mentor, with smaller values indicating greater competitiveness.

[c]<.001.

[d]<.05.

[e]<.01.

[f]SNIP: Source Normalized Impact per Paper.

[g]No results were produced by XLSTAT software for this group.

For the majority of participants (basic medicine, internal medicine, and surgery), becoming a high-funding recipient required early success. Specifically, higher SNIP_0–5y values were positively associated with becoming a high-funding recipient in basic medicine (odds ratio [OR] 1.48, 95% CI 1.22 - 1.79), internal medicine (OR 1.34, 95% CI 1.06 - 1.69), and surgery (OR 1.70, 95% CI 1.28 - 2.27). Likewise, researchers whose mentors had higher competitiveness scores were more likely to become high-funding recipients, as indicated by the inverse association with the mentor-level score (basic medicine: OR 0.49, 95% CI 0.36 - 0.66; internal medicine: OR 0.67, 95% CI 0.46 - 0.96; surgery: OR 0.38, 95% CI 0.22 - 0.68). Sustained high-quality leadership outputs were also significant (average SNIP: basic medicine, OR 5.99, 95% CI 2.63 - 13.65; internal medicine, OR 3.17, 95% CI 1.16 - 8.70; surgery, OR 9.79, 95% CI 3.26 - 29.43; Table 6). This was further supported by the results of our analysis that explored factors characterizing high-funding recipients without distinguishing fields (Multimedia Appendices 3 and 4).

For surgery, in addition to these factors, supporting specific collaborators (surgery: elite reliance: OR 2.09, 95% CI 1.10 - 3.97) and the quality of publications during the first 12 years of a researcher's career were major determinants of success (surgery: SNIP_first_0–12y, OR 1.56, 95% CI 1.04 - 2.35; SNIP_second_0–12y: OR 2.22, 95% CI 1.18 - 4.17; Table 6).

As funding acquisition data were characterized by small amounts for most researchers and large outliers for a limited number of researchers, some estimates showed wide CIs. Therefore, these associations should be regarded as exploratory. Nevertheless, positive associations (Multimedia Appendix 5) with high funding were observed for elite reliance in the field of social medicine (OR 9.09, 95% CI 1.32 - 62.57), early-career senior-author impact (SNIP_senior_0 - 12y: OR 11.66, 95% CI 1.75 - 77.73), and engagement in challenging research (% challenging research: OR 12.67, 95% CI 1.13 - 142.10; Table 6). In dentistry, selective mobilization capacity (OR 5.98, 95% CI 1.42 - 25.21) and late-career second-author impact (SNIP_second_25+y: OR 4.80, 95% CI 1.07 - 21.55) were also positively associated with high funding (Table 6).

To assess the robustness of funding acquisition patterns, we conducted multiple linear regressions using personal attributes and behavioral characteristics. As total funding raised (indicator: total funding) was inherently related to the total number of funded projects (indicator: # projects), the latter variable was excluded from the regression analyses to avoid structural redundancy. Consistent with this relationship, the 2 variables were strongly correlated ($r$=0.847; Multimedia Appendix 6).

To mitigate concerns regarding reverse causality and the influence of unmeasured variables, we first constructed models using theoretically relevant covariates for each career stage and assessed multicollinearity prior to model selection (Table 7). Model 1 included variables related to personal attributes and



XSL•FO
RenderX

early-career activities. In model 2, we additionally included early-career funding acquisition to account for the cumulative effects of prior funding success. Following the inclusion of this variable, the coefficient for graduate university rank changed direction but remained close to zero and statistically nonsignificant (from −0.02±0.04 to 0.02±0.03; not significant; Table 7), suggesting minimal independent association with subsequent funding acquisition. As the influence of graduate university rank appeared to be limited and largely accounted for by early-career funding success, this variable was not retained in subsequent models.



XSL•FO
RenderX

**Table .** Comparison of sequential regression model specifications for funding acquisition.[a]

| Metrics | Model 1 ($N_2$=804) | Model 2 ($N_2$=804) | Model 3 ($N_2$=804) | Model 4 ($N_2$=804) |
|---|---|---|---|---|
| Career length, standardized coefficient (SE) | 0.11 (0.03)[b] | 0.14 (0.03)[b] | 0.15 (0.03)[b] | 0.09 (0.03)[c] |
| University rank (currently affiliated), standardized coefficient (SE)[d] | −0.19 (0.04)[b] | −0.19 (0.03)[b] | −0.18 (0.03)[b] | −0.16 (0.03)[b] |
| University rank (graduated), standardized coefficient (SE)[d] | −0.02 (0.04) | 0.02 (0.03) | —[e] | — |
| Mentor level, standardized coefficient (SE)[f] | −0.27 (0.03)[b] | −0.24 (0.03)[b] | 0.12 (0.03)[b] | 0.12 (0.04)[b] |
| # first-authored_0–5y, standardized coefficient (SE) | −0.02 (0.03) | −0.02 (0.03) | −0.21 (0.03)[b] | −0.21 (0.03)[b] |
| SNIP_0–5y, standardized coefficient (SE) | 0.21 (0.04)[b] | 0.17 (0.03)[b] | −0.03 (0.03) | −0.04 (0.03) |
| Funding_0–12y, standardized coefficient (SE) | — | 0.32 (0.03)[b] | 0.32 (0.03)[b] | 0.32 (0.03)[b] |
| % lead-authored publications, standardized coefficient (SE) | — | — | 0.14 (0.03)[b] | 0.07 (0.03)[c] |
| Average SNIP, standardized coefficient (SE)[g] | — | — | 0.09 (0.03)[h] | — |
| SNIP_second_0 - 12y, standardized coefficient (SE) | — | — | — | 0.03 (0.03) |
| SNIP_second_13 - 24y, standardized coefficient (SE) | — | — | — | −0.02 (0.03) |
| SNIP_second_25+y, standardized coefficient (SE) | — | — | — | 0.04 (0.03) |
| SNIP_senior_0 - 12y, standardized coefficient (SE) | — | — | — | 0.12 (0.03)[b] |
| SNIP_senior_13 - 24y, standardized coefficient (SE) | — | — | — | 0.1 (0.03)[h] |
| SNIP_senior_25+y, standardized coefficient (SE) | — | — | — | 0.06 (0.03) |
| Male, standardized coefficient (SE) | −0.02 (0.03) | −0.03 (0.03) | −0.03 (0.03) | −0.03 (0.03) |
| Female, standardized coefficient (SE) | 0 (0) | 0 (0) | 0 (0) | 0 (0) |
| VIF[i] | 1.046 | 1.103 | 1.120 | 1.132 |
| *F* test (*df*) | 30.364 (7, 796) | 43.806 (8, 795) | 42.878 (9, 794) | 29.255 (14, 789) |
| Adjusted $R^2$ | 0.211 | 0.299 | 0.319 | 0.330 |
| AIC[j] | 16115.889 | 16014.542 | 15991.702 | 15984.003 |

[a]The dependent variable was the sum of the amount allocated for each project obtained as principal investigator.

[b]*P*<.001.

[c]*P*<.05.

[d]These variables were quantified based on The Times Higher Education World University Rankings 2023, with smaller values indicating higher ranking.

[e]Not applicable.

[f]These variables were quantified based on the largest grant category secured by the mentor, with smaller values indicating greater competitiveness.

[g]SNIP: Source Normalized Impact per Paper.



XSL•FO
RenderX

[h]$P<.01$.

[i]VIF: variance inflation factor.

[j]AIC: Akaike information criterion.

Model 3 further incorporated variables reflecting research activity across the entire career (indicators: % lead-authored publications and average SNIP). Following the inclusion of these variables, the coefficient for mentor level reversed direction while remaining statistically significant—it changed from −0.24±0.03 to 0.12±0.03 ($P<.001$). The coefficient for early success became significant for the number of papers (indicator: # first-authored_0–5y, −0.02±0.03 to −0.21±0.03; $P<.001$), whereas the coefficient for impact lost statistical significance (indicator: SNIP_0–5y: 0.17±0.03 to −0.03±0.03; $P>.05$). This suggests that part of the influence of early-career variables may be mediated through subsequent career experiences. To reflect the temporal structure of career development and clarify how associations change across career stages, we decided to retain the early-career variables in the model. In model 3, variables related to collaborative research relationships and acquisition strategies were excluded (indicators: elite reliance, selective mobilization capacity, # large-scale projects, and % challenging research; Table 7) because their univariate associations demonstrated wide CIs and appeared to be strongly influenced by extreme outliers (Table 6).

Finally, model 4 replaced the career-wide research quality variable used in model 3 (indicator: average SNIP) with research quality variables stratified by career stage and author position within 12-year career intervals (Table 7). As model 1 already captured early-career activity patterns as a first-author during the first 5 years (indicator: # first-authored_0–5y and SNIP_0–5y), additional variables representing first-author activity across the first 12 career years were not included in model 4 to avoid redundancy among temporally overlapping indicators. Furthermore, first-author activity indicators from the midcareer stage onward were excluded because their univariate associations were not statistically significant (Table 6). Therefore, only variables representing second-author and senior-author activity from the midcareer stage onward were added.

After confirming acceptable variance inflation factor values across alternative model specifications (Table 7), model 4 was selected as the final modeling framework because it allowed stage-specific research activity indicators to be evaluated while minimizing redundancy among temporally overlapping variables. Field-specific regression models were subsequently constructed based on this framework (Table 8). The model fit, evaluated with adjusted $R^2$ values, was high for social medicine and dentistry with low Akaike information criterion (AIC) values, suggesting that funding acquisition in these fields can be explained and predicted with considerable accuracy (model 4: adjusted $R^2$=0.890; model 6: adjusted $R^2$=0.738; Table 8). In surgery, an adjusted $R^2$ of 0.576 indicated a reasonably good predictive performance, although a high AIC suggested limited justification for retaining variables with nonsignificant coefficients. Conversely, the results for nurses showed limited explanatory power, suggesting that key determinants in this field might not have been fully captured (model 5: $F$ test not significant; adjusted $R^2$=0.047; Table 8). For basic medicine and internal medicine, relatively low adjusted $R^2$ values and high AIC values suggested that additional factors outside this analysis may influence funding acquisition (model 1: adjusted $R^2$=0.288; model 2: adjusted $R^2$=0.318; Table 8).



XSL•FO
RenderX

**Table .** Probability of funding acquisition by field (based on a 2-tailed *t* test; ordinary least squares was used for analysis)[a].

| Metrics | Field | | | | | |
|---|---|---|---|---|---|---|
| | Basic medicine (n=240) | Clinical medicine (internal medicine; n=162) | Clinical medicine (surgery; n=247) | Social medicine (n=48) | Nursing (n=54) | Dentistry (n=53) |
| Career length, standardized coefficient (SE) | 0.10 (0.07) | 0.09 (0.08) | 0.16 (0.05)[b] | 0.20 (0.08)[c] | 0.22 (0.19) | 0.02 (0.11) |
| University rank (currently affiliated), standardized coefficient (SE)[d] | 0.21 (0.06)[e] | 0.38 (0.07)[e] | 0.52 (0.04)[e] | 0.91 (0.07)[e] | 0.03 (0.16) | 0.39 (0.09)[e] |
| Mentor level, standardized coefficient (SE)[f] | −0.25 (0.06)[e] | −0.24 (0.07)[e] | −0.07(0.04) | −0.07 (0.06) | −0.49 (0.2) | −0.03 (0.08) |
| # first-authored_0–5y, standardized coefficient (SE) | −0.19 (0.06)[b] | −0.15 (0.07)[c] | −0.34 (0.05)[e] | 0.01 (0.06) | −0.20 (0.15) | 0.04 (0.09) |
| SNIP_0–5y, standardized coefficient (SE) | −0.06 (0.06) | 0.02 (0.08) | −0.07 (0.05) | 0.00 (0.06) | 0.16 (0.24) | −0.22 (0.09)[c] |
| Funding_0–12y, standardized coefficient (SE) | 0.11 (0.07) | −0.05 (0.08) | 0.20 (0.05)[e] | −0.01 (0.08) | −0.12 (0.28) | 0.46 (0.11)[e] |
| % lead-authored publications, standardized coefficient (SE) | 0.08 (0.06) | 0.11 (0.07) | 0.01 (0.05) | 0.06 (0.06) | −0.02 (0.15) | 0.04 (0.09) |
| SNIP_second_0 - 12y, standardized coefficient (SE) | 0.02 (0.06) | 0.05 (0.07) | 0.03 (0.05) | 0.15 (0.07)[c] | −0.02 (0.18) | 0.39 (0.1)[e] |
| SNIP_second_13 - 24y, standardized coefficient (SE) | 0.03 (0.06) | −0.05 (0.07) | −0.04 (0.05) | −0.10 (0.07) | −0.21 (0.19) | 0.24 (0.1)[c] |
| SNIP_second_25+y, standardized coefficient (SE) | 0.03 (0.06) | 0.07 (0.07) | 0.07 (0.05) | 0.04 (0.07) | −0.05 (0.17) | 0.17 (0.09) |
| SNIP_senior_0 - 12y, standardized coefficient (SE) | 0.20 (0.06)[e] | 0.11 (0.07) | 0.02 (0.05) | −0.04 (0.07) | 0.17 (0.17) | 0.05 (0.09) |
| SNIP_senior_13 - 24y, standardized coefficient (SE) | 0.04 (0.06) | 0.24 (0.08)[b] | 0.06 (0.05) | 0.20 (0.06)[b] | −0.16 (0.16) | −0.02 (0.1) |
| SNIP_senior_25+y, standardized coefficient (SE) | 0.04 (0.06) | 0.10 (0.07) | 0.04 (0.05) | −0.01 (0.06) | −0.08 (0.15) | −0.14 (0.08) |
| Male, standardized coefficient (SE) | −0.06 (0.06) | 0.05 (0.07) | −0.01 (0.04) | −0.03 (0.05) | −0.11 (0.18) | 0.03 (0.09) |
| Female, standardized coefficient (SE) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) |
| *F* test (*df*) | 7.902 (14, 225)[e] | 6.368 (14, 147)[e] | 24.910 (14, 232)[e] | 28.280 (14, 33)[e] | 0.322 (14, 39) | 11.441 (15, 38)[e] |
| Adjusted $R^2$ | 0.288 | 0.318 | 0.576 | 0.890 | 0.047 | 0.738 |
| AIC[g] | 4976.318 | 3217.847 | 4371.011 | 876.836 | 932.875 | 919.093 |



XSL•FO
RenderX

[a]The dependent variable was the sum of the amount allocated for each project obtained as principal investigator.

[b]$P<.01$.

[c]$P<.05$.

[d]These variables were quantified based on The Times Higher Education World University Rankings 2023, with smaller values indicating higher ranking.

[e]$P<.001$.

[f]These variables were quantified based on the largest grant category secured by the mentor, with smaller values indicating greater competitiveness.

[g]AIC: Akaike information criterion.

## Discussion

### Principal Findings

Funding disparities across academic fields have been analyzed through publication-linked funding information, showing both the breadth and concentration of funding distribution within medical subfields [25,26]. However, because these are research evaluations rather than assessments of individual achievements, they do not sufficiently analyze the differences in researchers' abilities to secure funding. The analysis demonstrated that the maximum funding secured by individual researchers varies by field.

One possible explanation for the limited number of high-funding recipients in nursing is the study design itself. While the initial data extraction yielded a sample believed to reflect the structure of the population, subsequent procedures—specifically the inclusion criteria of being affiliated with a national university medical school and having obtained a doctoral degree since 1996—excluded researchers from private universities with well-established nursing programs, as well as many experienced nursing researchers whose scholarship is grounded primarily in extensive clinical practice (from 9.96% to 5.12%; Table 1). On the other hand, observations from researchers affiliated with Japanese medical schools suggest that highly funded nursing researchers are often leaders who have created and are driving emerging subfields such as patient-participatory nursing and nursing digital transformation. Consequently, the research population may be smaller in these new subfields. This may partly explain the limited number of recipients of very large-scale research funding in nursing.

The scarcity of high-funding recipients in surgery and dentistry (Table 4) may reflect the historical origins of these practice-oriented professions and their field cultures that traditionally value technical crafts. Clinically oriented publications receive fewer citations than those focused on basic science, making funding acquisition more difficult [27]. While low citation counts may reflect a narrow readership due to high specialization, there is evidence that lower-prestige publications may hinder the diffusion of cross-disciplinary ideas [28]. US NIH data indicate that selecting themes favored by a broader scientific community (ie, more fundamental or mechanistic themes) can influence funding decisions regardless of the quality of future outcomes [29]. In Japan's GIA large-scale projects, the broader range of reviewers' specialties compared with small-scale projects may be one reason fewer high-funding recipients are found in surgery and dentistry. Greater clinical demands, driven by the growing importance of revenue-generating activities for university hospitals following the 2004 partial privatization reform of national universities, may also contribute [30].

This study showed the impact of individual efforts, such as demonstrating leadership to produce high-quality outcomes at each career stage (Table 6). The importance of second authorship, previously underexamined, indicates the degree of centralization of the research ecosystem in each field [31]. The finding that early success in surgery accounted for nearly the entire outcome (Table 8) suggests a strong influence of institutional and academic structures. However, the reversal of the coefficients for mentor level and early success after controlling for career-long achievements suggests that the apparent effects of these early-career factors may be partly attributable to their association with subsequent research performance (Table 7). Funding success therefore appears to reflect not only early-career advantages but also achievements accumulated throughout a research career.

Promoting complementary research while supporting specific collaborators in social medicine or maintaining a subleadership position while mobilizing specific researchers as supporters in dentistry (Table 6) can be regarded as evidence that researchers are taking proactive measures in these fields. The field-specific research design of this study, by excluding typical leaders who tend to attract an excessive concentration of funding in these fields, may have highlighted the importance of the actions of individuals with relatively less influence. Examining the moderate funding recipients and diverse research behaviors may be more informative than focusing solely on top recipients.

This study provides new insights into individual agency within the process through which researchers transition from peripheral to central roles. As effective individual actions and timing for securing funding vary by field, it is necessary to concentrate efforts at the most advantageous time (Table 6).

### Limitations

The limitations of this study are as follows. First, it does not cover all professors conducting research with GIA grants, as the analysis is restricted to researchers who obtained their doctoral degrees after 1996 and whose publication records can be traced from the early stages of their careers. Consequently, researchers in the oldest age group and those without doctoral degrees are excluded. However, this restriction is consistent with the purpose of the study. As Japan's research funding system shifted toward a competitive funding model in 2004, focusing on researchers whose careers developed under this system provides an appropriate basis for examining its effects. Accordingly, the findings should be interpreted as reflecting trends within the post-2004, globally standardized research environment. The experiences of older researchers and those without doctoral degrees remain important topics for future



XSL•FO
RenderX

research. In particular, historical comparisons could provide valuable insights into how research was nurtured within Japan's pre-2004 research ecosystem.

Second, SciVal is a commercial database with limited journal coverage, and reliance on this source alone may underestimate researchers' publication output. The extent of this underestimation may vary across fields. On the basis of our previous analysis of researcher career outcomes, SciVal data appear to reasonably capture researcher performance in fields with globally shared research interests, such as basic and clinical medicine [32]. In contrast, the limitations are more pronounced in fields where findings are frequently disseminated in local languages. Further discipline-specific investigations are needed. Furthermore, this analysis included all forms of research output to capture diverse scholarly contributions, including conceptual synthesis in review articles. While our preliminary analyses suggest that review articles are not strongly associated with grant acquisition, recent advances in AI and laboratory automation suggest the need to reevaluate scholarly contributions beyond traditional academic publications.

The third limitation concerns the indicators used in this study. While this study demonstrated that second authorship can serve as a stepping stone to a core role, the path to senior author excellence remains unclear. Future research should develop new metrics that capture the diverse contributions of middle authors to understand the determinants of core and peripheral roles [33,34]. Moreover, it could not account for aspirations to lead large-scale research projects, which may be partially reflected in their rates of submitting large-scale grant applications. Future analyses should incorporate variations in application behavior to better capture such ambitions. In addition, funding acquisition may also be influenced by finer behavioral differences, such as promotional language in proposals. Addressing these factors will require new behavioral measures beyond existing research evaluation metrics.

Finally, potential confounding between early-career variables and total-career indices should be considered when interpreting the results. Early-career variables were retained in the models because the attenuation of their effects after adjustment for total-career factors was itself a finding of interest. As a result, some confounding may remain, and the estimated coefficients should be interpreted with caution. Furthermore, residual confounding due to unmeasured factors cannot be excluded, limiting causal interpretation of the observed associations.

## Conclusions

This study identifies individual research-related behaviors that influence GIA funding acquisition across research fields. The results highlight that while organizational and academic structures have a significant impact, personal initiative during researchers' transition from peripheral to core roles becomes important, emphasizing strong relationship-building skills. In other words, both participants and coordinators in collaborative research, commitment to reshaping research agendas, and excellent publication records as the lead author are crucial factors that affect funding acquisition. Effective actions and timing vary by field, reflecting the maturity of each field's ecosystem. These survey results indicate the actions and timing that enable individual researchers to maximize funding acquisition and expand their research, providing a foundation for more effective institutional support strategies.

## Acknowledgments

The authors thank Naoko Manabe, Natsuho Ishikawa, and Himika Tamura for assistance with data acquisition. During the preparation of this paper, we used generative AI to improve its readability. Furthermore, the text was edited by a professional English proofreader. The authors thank Editage [35] for English language editing.

## Funding

This study was supported by grants from the Japan Society for the Promotion of Science (JSPS KAKENHI grant 23K02545), Sumitomo Electric Group, and Inamori Research Grants to AH.

## Data Availability

The datasets generated or analyzed during this study are not publicly available because they contain publication data obtained from commercial databases but are available from the corresponding author on reasonable request.

## Authors' Contributions

AH contributed to conceptualization, data curation, investigation, visualization, and original draft preparation. ST contributed to conceptualization, resources, writing, reviewing, and editing.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
The metrics.
[PDF File, 346 KB - i-jmr_v15i1e98428_app1.pdf ]



XSL•FO
RenderX

Multimedia Appendix 2
Phenotype clusters.
[PDF File, 387 KB - i-jmr_v15i1e98428_app2.pdf ]

Multimedia Appendix 3
Behavioral characteristics by phenotype clusters.
[PDF File, 684 KB - i-jmr_v15i1e98428_app3.pdf ]

Multimedia Appendix 4
Associations between behavioral characteristics and funding acquisition by phenotype clusters.
[PDF File, 943 KB - i-jmr_v15i1e98428_app4.pdf ]

Multimedia Appendix 5
Regression of behavior-related indicators against the amount allocated (log-transformed).
[PDF File, 248 KB - i-jmr_v15i1e98428_app5.pdf ]

Multimedia Appendix 6
Correlation matrix.
[PDF File, 788 KB - i-jmr_v15i1e98428_app6.pdf ]

## References


1. Lim BH, D'Ippoliti C, Dominik M, et al. Regional and institutional trends in assessment for academic promotion. Nature 2025 Feb;638(8050):459-468. [doi: 10.1038/s41586-024-08422-9] [Medline: 39843736]
2. Badelt C. Private external funding of universities: blind alley or new opening? Rev Manag Sci 2020 Apr;14(2):447-458. [doi: 10.1007/s11846-019-00365-0]
3. Bol T, de Vaan M, van de Rijt A. The Matthew effect in science funding. Proc Natl Acad Sci U S A 2018 May 8;115(19):4887-4890. [doi: 10.1073/pnas.1719557115] [Medline: 29686094]
4. van den Besselaar P, Sandström U. Early career grants, performance, and careers: a study on predictive validity of grant decisions. J Informetr 2015 Oct;9(4):826-838. [doi: 10.1016/j.joi.2015.07.011]
5. van Arensbergen P, van der Weijden I, van den Besselaar P. Different views on scholarly talent: what are the talents we are looking for in science? Res Eval 2014 Oct 1;23(4):273-284. [doi: 10.1093/reseval/rvu015]
6. Ma A, Mondragón RJ, Latora V. Anatomy of funded research in science. Proc Natl Acad Sci U S A 2015 Dec 1;112(48):14760-14765. [doi: 10.1073/pnas.1513651112] [Medline: 26504240]
7. Lauer MS, Roychowdhury D. Inequalities in the distribution of National Institutes of Health research project grant funding. Elife 2021 Sep 3;10:e71712. [doi: 10.7554/eLife.71712] [Medline: 34477108]
8. Liao CH. The Matthew effect and the halo effect in research funding. J Informetr 2021 Feb;15(1):101108. [doi: 10.1016/j.joi.2020.101108]
9. van der Lee R, Ellemers N. Gender contributes to personal research funding success in The Netherlands. Proc Natl Acad Sci USA 2015 Oct 6;112(40):12349-12353. [doi: 10.1073/pnas.1510159112]
10. Rusu VD, Mocanu M, Bibiri AD. Determining factors of participation and success rates in research funding competitions: case study. PLoS One 2022;17(7):e0272292. [doi: 10.1371/journal.pone.0272292] [Medline: 35905079]
11. Mocanu M, Rusu VD, Bibiri AD. Competing for research funding: key elements impacting the evaluation of grant proposal. Heliyon 2024 Aug 30;10(16):e36015. [doi: 10.1016/j.heliyon.2024.e36015] [Medline: 39224302]
12. Tsugawa S, Kanetsuki T, Sugihara J. Relationship between early-career collaboration among researchers and future funding success in Japanese academia. PLoS One 2022;17(11):e0277621. [doi: 10.1371/journal.pone.0277621] [Medline: 36367875]
13. Liu X, Ou G, Chen C. Good to great: the impact of interdisciplinarity on the researchers' funding performance. Data and Information Management 2025 Sep;9(3):100094. [doi: 10.1016/j.dim.2025.100094]
14. Deville P, Wang D, Sinatra R, Song C, Blondel VD, Barabási AL. Career on the move: geography, stratification, and scientific impact. Sci Rep 2014 Apr 24;4(4):4770. [doi: 10.1038/srep04770] [Medline: 24759743]
15. Gonzalez-Brambila CN, Veloso FM, Krackhardt D. The impact of network embeddedness on research output. Res Policy 2013 Nov;42(9):1555-1567. [doi: 10.1016/j.respol.2013.07.008]
16. Mudrak J, Zabrodska K, Machovcova K. Crafting cumulative advantage: a systemic approach to the career development of highly productive researchers. Minerva 2026 Mar;64(1):191-216. [doi: 10.1007/s11024-024-09567-7]
17. Alisic A, Wiese BS. Keeping an insecure career under control: the longitudinal interplay of career insecurity, self-management, and self-efficacy. J Vocat Behav 2020 Aug;120:103431. [doi: 10.1016/j.jvb.2020.103431]
18. Zaorsky NG, O'Brien E, Mardini J, Lehrer EJ, Holliday E, Weisman CS. Publication productivity and academic rank in medicine: a systematic review and meta-analysis. Acad Med 2020 Aug;95(8):1274-1282. [doi: 10.1097/ACM.0000000000003185] [Medline: 32028299]




XSL•FO
RenderX

19. Nguyen M, Panyadahundi A, Olagun-Samuel C, et al. Transition from mentored to independent NIH funding by gender and department. JAMA 2023 Jun 27;329(24):2189-2190. [doi: 10.1001/jama.2023.7693] [Medline: 37367985]
20. Search for researchers. Kaken. URL: https://nrid.nii.ac.jp/ja/index/ [accessed 2026-02-26]
21. CiNii Research. CiNii. URL: https://ci.nii.ac.jp/ [accessed 2026-02-26]
22. SciVal. URL: https://www.scival.com [accessed 2026-02-26]
23. Hashiguchi A, Asashima M, Takahashi S. The influence of human connections and collaboration on research grant success at various career stages: regression analysis. JMIR Form Res 2024 Feb 28;8:e49905. [doi: 10.2196/49905] [Medline: 38416548]
24. Hashiguchi A, Watanabe KN, Takahashi S. Assessing the impact of Japanese team grants in plant science. Plant Cell Physiol 2026 Feb 27;67(2):111-114. [doi: 10.1093/pcp/pcaf153] [Medline: 41269269]
25. Tian W, Cai R, Fang Z, Xie Q, Hu Z, Wang X. Research funding in different SCI disciplines: a comparison analysis based on Web of Science. Quantitative Science Studies 2024 Aug 1;5(3):757-777. [doi: 10.1162/qss_a_00315]
26. Thelwall M, Kousha K, Abdoli M, et al. Is research funding always beneficial? A cross-disciplinary analysis of U.K. research 2014–20. Quantitative Science Studies 2023 May 1;4(2):501-534. [doi: 10.1162/qss_a_00254]
27. Ke Q. The citation disadvantage of clinical research. J Informetr 2020 Feb;14(1):100998. [doi: 10.1016/j.joi.2019.100998]
28. Morgan AC, Economou DJ, Way SF, Clauset A. Prestige drives epistemic inequality in the diffusion of scientific ideas. EPJ Data Sci 2018 Dec;7(1):40. [doi: 10.1140/epjds/s13688-018-0166-4]
29. Hoppe TA, Litovitz A, Willis KA, et al. Topic choice contributes to the lower rate of NIH awards to African-American/black scientists. Sci Adv 2019 Oct;5(10):eaaw7238. [doi: 10.1126/sciadv.aaw7238] [Medline: 31633016]
30. Kikuchi Y. Impact of university reform on research performance aggregated and disaggregated across research fields: a case study of the partial privatization of Japanese national universities. JER 2023 Jan;74(1):1-27. [doi: 10.1007/s42973-021-00074-y]
31. Michelson KA. Association of publication record and independent NIH funding. PLoS One 2022;17(6):e0269283. [doi: 10.1371/journal.pone.0269283] [Medline: 35771742]
32. Hashiguchi A, Hamada H, Takahashi S, Honda K. Impact of supervisors' research style on young biomedical scientists' capacity development as measured by REDi, a novel index of crossdisciplinarity. Front Res Metr Anal 2022;7:990921. [doi: 10.3389/frma.2022.990921] [Medline: 36204454]
33. Sundling P. Author contributions and allocation of authorship credit: testing the validity of different counting methods in the field of chemical biology. Scientometrics 2023 May;128(5):2737-2762. [doi: 10.1007/s11192-023-04680-y]
34. Yang S, Xiao A, Nie Y, Dong J. Measuring coauthors' credit in medicine field — Based on author contribution statement and citation context analysis. Information Processing & Management 2022 May;59(3):102924. [doi: 10.1016/j.ipm.2022.102924]
35. Editage. URL: https://www.editage.com/ [accessed 2026-07-07]


## Abbreviations

**AIC:** Akaike information criterion
**GIA:** Grant-in-Aid for Scientific Research
**NIH:** National Institutes of Health
**OR:** odds ratio
**PI:** principal investigator
**SNIP:** Source Normalized Impact per Paper


*Edited by M Balcarras; submitted 15.Apr.2026; peer-reviewed by R Lordan, Y Jiang; revised version received 08.Jun.2026; accepted 23.Jun.2026; published 27.Jul.2026.*

*Please cite as:*
*Hashiguchi A, Takahashi S*
*Individual Research Behaviors and Research Funding Acquisition Across Fields and Career Periods: Regression Analysis*
*Interact J Med Res 2026;15:e98428*
*URL: https://www.i-jmr.org/2026/1/e98428*
*doi:10.2196/98428*


© Akiko Hashiguchi, Satoru Takahashi. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 27.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete



XSL•FO
RenderX

bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Innovation Deimplementation in Emergency Departments During the COVID-19 Pandemic: Qualitative Study of Clinicians' Decision-Making

Shreya Huilgol[1,2], BA; Nabeel Qureshi[2,3], MPH, MPhil, PhD; Carl Berdahl[2,3], MS, MD; Catherine Cohen[2], RN, PhD; Peter Mendel[2], PhD; Shira Fischer[4], MD, PhD

[1]Department of Health Policy and Management, University of California, Los Angeles, 650 Charles E Young Dr S, Los Angeles, CA, United States

[2]RAND, Santa Monica, CA, United States

[3]Cedars Sinai Medical Center, Los Angeles, CA, United States

[4]RAND, Boston, MA, United States

**Corresponding Author:**
Shreya Huilgol, BA
Department of Health Policy and Management, University of California, Los Angeles, 650 Charles E Young Dr S, Los Angeles, CA, United States

## Abstract

**Background:** During a public health emergency, emergency department (ED) clinicians can improve care delivery if they identify and adopt innovations that are safe and effective. However, little is known about the factors that impact ED clinicians' decision-making around using or discontinuing innovations when evidence-based information is limited.

**Objective:** The goal of this study was to understand the processes and factors that led ED clinicians to discontinue (deimplement) the use of COVID-19 care innovations.

**Methods:** This is a qualitative study using semistructured focus groups with ED clinicians from 8 hospitals across the United States. Hospitals were purposively sampled and recruited to capture a diversity of perspectives based on location, facility type (academic or community hospital), rurality (urban or rural), and safety-net status. In this study, 17 physicians, 7 advanced practice providers, 18 nurses, and 7 respiratory therapists participated. We utilized both inductive and deductive techniques to perform content and thematic analysis of transcripts.

**Results:** Clinicians shared that their own experiences (eg, direct observation of patient outcomes), contextual factors, and emerging research evidence contributed heavily to decisions about deimplementing innovations during the COVID-19 pandemic. Processes related to discontinuing innovations depended on leadership guidance and collaboration among colleagues. However, in some cases, there were no official processes to discontinue innovations, and innovations were passively deimplemented.

**Conclusions:** Decision-making regarding the discontinuation of innovation in ED settings during the COVID-19 pandemic differed from routine conditions due to the lack of information and the rapid evolution of evidence within a short period of time. The level of evidence required to implement and deimplement innovations was significantly lower. Our findings indicate that factors influencing deimplementation during a public health emergency were highly localized and were treated similarly to pilot tests of new innovations. Future work is necessary to develop mechanisms for implementing promising innovations during evolving public health emergencies and monitoring their effectiveness and safety after implementation, enabling evidence-based decisions about whether to continue implementation or proceed with deimplementation.

***(Interact J Med Res 2026;15:e82088)*** doi:10.2196/82088

**KEYWORDS**

deimplementation; COVID-19 pandemic; innovation; emergency department; decision-making

## Introduction

During the COVID-19 pandemic, emergency department (ED) clinicians influenced patient health outcomes by rapidly identifying and implementing care innovations to manage and treat patients with COVID-19. We define care innovations as "new or improved health policies, systems, products and technologies, and services and delivery methods that improve people's health." [1] Examples of care innovations used in EDs during COVID-19 included utilizing prone positioning [2], high-flow oxygen [3], telemedicine [4,5], and intubation boxes [6], as well as reusing personal protective equipment (PPE) [7].

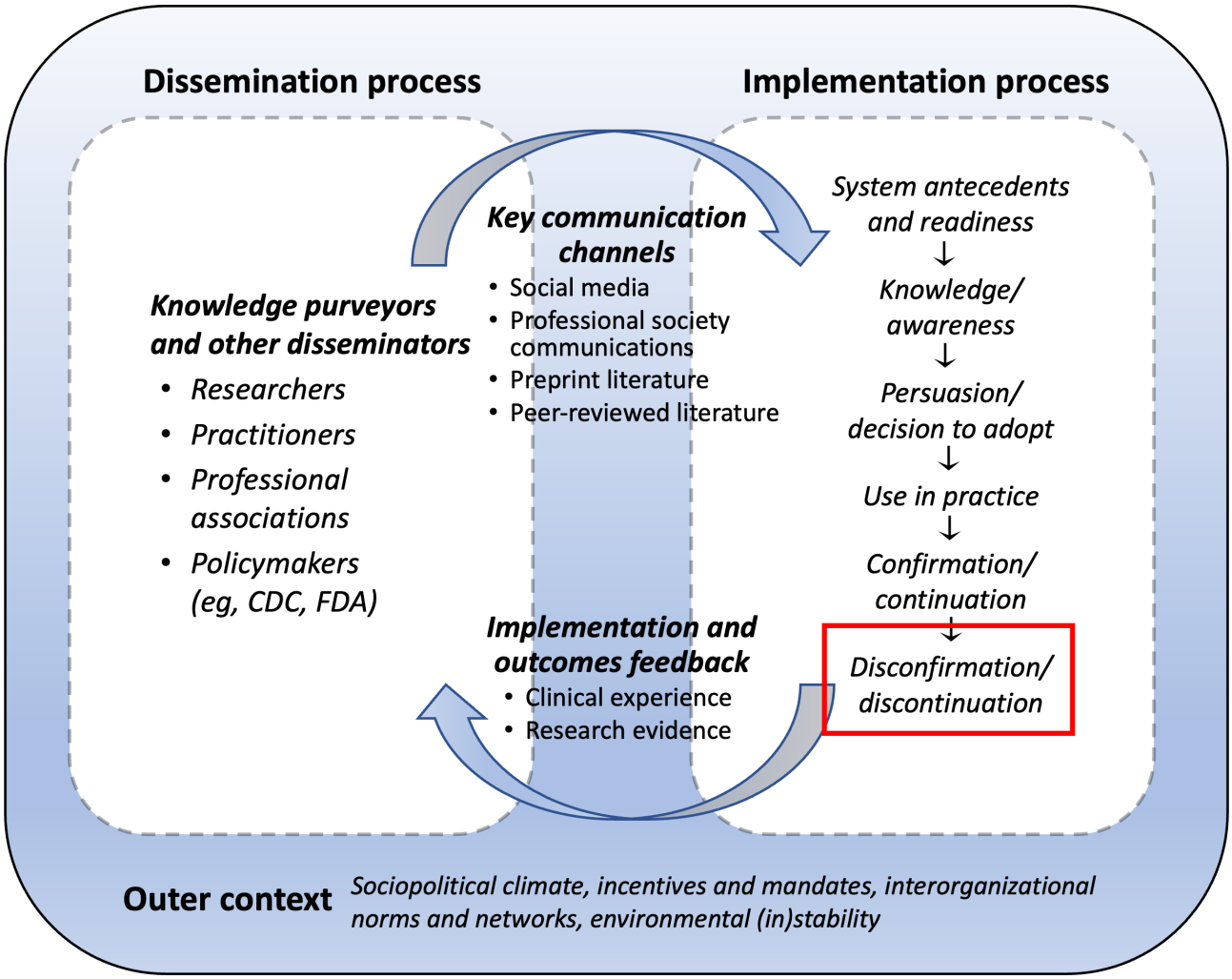
Amid the public health emergency, ED clinicians learned about innovations (dissemination process) and decided whether and



XSL•FO
RenderX

how to implement them (implementation process; see Figure 1). Many innovations were also adopted throughout the hospital in response to COVID-19, but a large number were first implemented in the ED. As new evidence emerged, there were changes in innovation adoption and use: some innovations were sustained, while others were discontinued. In response to emerging anecdotal and peer-reviewed evidence during the pandemic, clinicians discontinued practices that proved unnecessary, ineffective, or even dangerous. This process is called deimplementation. Deimplementation, or deadoption, is the process by which practices deemed harmful, ineffective, or no longer necessary are discontinued within an organization [8-10].

**Figure 1.** Diffusion framework for COVID-19-related emergency department innovations [11]. CDC: Centers for Disease Control and Prevention; FDA: US Food and Drug Administration.




While the implementation of care innovations has been and continues to be well studied in the ED setting [12,13], less attention has been paid to deimplementation. Deimplementation differs from implementation in many important ways. A recent review noted that psychological factors and biases are more relevant to deimplementation than implementation, as many deimplementation activities involve stakeholders actively changing their behaviors [14,15]. Consequently, different behavioral change techniques are used in implementation and deimplementation studies, making the case that deimplementation should be studied as a separate process from implementation [16]. Compared to implementation, deimplementation seems to be hindered more greatly by local (ie, departmental) priorities and biases among leaders and practitioners within health systems, in addition to the traditional outside factors impacting implementation, such as social and economic barriers [14], likely due to the relative autonomy providers have in providing health services. Additionally, a narrative review of recent studies of embedded pragmatic clinical trials found that factors outside of effectiveness can determine deimplementation, such as perceived benefits, costs, and alignment with other policies or priorities [17]. However, information regarding factors that impact clinician decision-making around the deimplementation of innovations is limited [18], and even less is known about factors influencing deimplementation during a public health emergency, such as the COVID-19 pandemic. Further study of important processes and determinants for successful deimplementation is critical [19].

Attempts to formalize deimplementation in the context of implementation have linked both using 4 types of change for deimplementation: partial reduction, complete reversal, substitution with related replacement, and substitution with unrelated replacement of existing practices [20]. Unlike many



XSL•FO
RenderX

previous deimplementation studies, these changes are usually related to deimplementing care processes rather than reducing the use of low-value treatments, such as improper medication delivery. In the context of the COVID-19 pandemic, this provides an opportunity to update and enhance preexisting deimplementation literature to focus on cases where the deimplementation involves the entire intervention, not just the degree or dose [21]. The synthesis model proposed by Niven et al [22] for the process of deadoption better aligns with deimplementation decision-making in the ED: identifying and prioritizing low-value practices is a key component of the model.

This study is part of a larger examination of the dissemination and implementation of care innovations during COVID-19. We utilized an adapted form of the Greenhalgh et al [11] framework for the diffusion of innovations in health services to systematically assess both implementation and deimplementation activities. In this paper, we examined factors relating to the deimplementation of innovations and the resulting deimplementation processes in the ED during the COVID-19 pandemic. The goal of this study was to understand how ED clinicians decided to discontinue using COVID-19 care innovations by identifying processes and factors that influenced decision-making.

## Methods

### Study Design and Setting

We conducted virtual pilot interviews and focus groups with ED clinicians to understand their experiences with practices related to the deimplementation of COVID-19 care innovations. The discussion guide was informed by a framework adapted from Greenhalgh et al [11] and modified after piloting with 1 physician, 1 nurse, and 1 nurse practitioner. After the initial pilot interviews, no changes were made to the content discussed or focus group guides. Thus, findings from the 3 pilot interviews were included in our overall analysis.

To meet our study objective, we conducted 13 subsequent focus group discussions with clinicians from 8 hospital-based EDs across the United States in 2022 and 2023. Because we believed that responses would differ by provider type and that professional hierarchy power dynamics could impact participant responses and comfort levels, we organized separate groups for ED prescribers (ie, physicians [medical doctors; MDs] and advanced practice providers [APPs]) and for other ED clinicians (ie, registered nurses [RNs] and respiratory therapists [RTs]) at each hospital to facilitate deeper conversations and limit concerns about nonparticipation [23].

### Focus Group Recruitment

We used maximum diversity sampling to recruit personnel in hospital–based EDs to participate in focus groups [24]. Hospitals were diverse by US Census Bureau region (4 regions) [25], academic hospital status (binary), rurality (binary), and safety net status (binary), including one at the epicenter of the pandemic (New York City). A site-specific champion was recruited by email and assisted with recruitment and scheduling.

### Participant Recruitment and Data Collection

Focus group participants were recruited by email. Participants were eligible if they had experience providing care in a participating hospital's ED during the COVID-19 pandemic. Forty-nine clinicians from 8 hospitals participated in the focus groups.

Focus groups lasted approximately 60 to 90 minutes and included at least 1 moderator and 1 notetaker. Three researchers (SF, CB, and CC) with both clinical and focus group experiences led the discussions, while other researchers (SH and NQ) took notes. Before each focus group, the moderator introduced themselves and provided a brief description of the research project. Focus group discussion topics included gaining information about COVID-19 treatment and sources of information, facilitators and barriers to adopting innovations, processes for confirming and sustaining promising innovations, processes for discontinuing innovations that were not promising, and how contextual factors impacted innovation implementation and deimplementation. Textbox 1 displays focus group questions specifically related to deimplementation, as discussed in this paper (for the full interview protocol, see Huilgol et al [26]). We continued conducting focus groups until thematic saturation was determined [27]. Study leaders assessed thematic saturation in team discussions after each focus group based on insights shared during the discussions.

**Textbox 1.** Sample focus group questions.

Deimplementing innovations

Sample questions:

- Tell me about a time you tried something and immediately said, I'm not doing that again? What was that like? What made you decide not to try it again?
- Here are some examples of things you might have stopped using: intubation boxes, reusing PPE, plasma, nasal steroids. Were any of these things you were doing and then stopped? Tell us about it. What made you make that decision?
- Can you give an example of one innovation you or your institution adopted and later rejected? Please describe how the decision was made to reject it. [Reject is defined as to stop using after having tried it.]
    - How did you personally decide whether to reject it?
    - How did your department or ED decide whether to reject it?
    - How did the institution as a whole decide to reject it?



XSL•FO
RenderX

### Ethical Considerations

This work was approved by the RAND Human Subjects Protection Committee (project 2021-N0714). Participants provided verbal consent at the start of the focus group and were compensated with a US $150 gift card for their time and participation. After receiving participant consent, focus groups were audio-recorded and transcribed for analysis. Participants' privacy and confidentiality were protected by securely storing focus group transcripts and recordings, limiting access to authorized researchers, and de-identifying data prior to analysis to ensure no personal information was linked to participant responses.

### Data Analysis

We used both deductive and inductive analysis techniques to thematically code the focus group data [28]. We started by creating a preliminary codebook defined by topics included in the focus group discussion guide. We then used inductive processes to identify broader patterns and themes from the coded data. After the first focus group was completed, 2 researchers with expertise in health services and qualitative methodology (SH and NQ) independently reviewed a single transcript and applied codes to assess the quality of the codebook and the frequency of codes, representing approximately 10% of all coded data. The 2 coders performed an interrater reliability test and calculated a κ score of 0.72, indicating good agreement [29]. The codebook was updated and finalized based on this coding. As new themes and insights emerged during focus group discussions, the authors updated and added new codes to the codebook. Additions to the codebook were shared via regular weekly meetings among the project team [30].

The remaining transcripts were then divided in half and each coded by a single author. The same 2 authors met regularly to review coding and address any questions that emerged to ensure consistent code application. After coding was completed, the authors discussed coded excerpts holistically to identify general themes and subthemes that differed across groups. To ensure accuracy, one author reviewed a transcript coded by the other author at the beginning and end of the coding process.

The results are categorized by the first 2 phases of deadoption from the framework proposed by Niven: (1) identifying and prioritizing low-value clinical practices and (2) facilitating the deimplementation process. These 2 phases complemented the Greenhalgh framework guiding our analysis as they further explicate specific issues related to deimplementation [11]. All qualitative analyses were conducted in Dedoose [31]. We followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) (Checklist 1) as a framework for data reporting [32]. Results about the implementation of innovations and information uptake in EDs have been published in detail elsewhere (see Huilgol et al [26] and Qureshi et al [33]).

## Results

### Hospital and Participant Characteristics

In addition to 3 pilot interviews, we conducted 13 semistructured focus group discussions. In total, 49 clinicians from 8 hospitals participated: 24 *prescribers* (17 ED physicians and 7 APPs) in a total of 7 focus groups, ranging from 3 to 6 participants per site focus group, and 25 *other clinicians* (18 RNs and 7 RTs) in a total of 6 focus groups, ranging from 3 to 6 participants per site focus group (Table 1). Fourteen additional clinicians were invited to participate in focus groups but declined, bringing our total response rate to 78%.

**Table .** Hospital and participant characteristics.

| Hospital ID | Region | Focus group type | | | | Total |
|---|---|---|---|---|---|---|
| | | Physicians (MDs[a]) | APPs[b] | RNs[c] | RTs[d] | |
| Hospital 1 | Midwest | 6 | 0 | 5 | 0 | 11 |
| Hospital 2 | West | 4 | 0 | 2 | 1 | 7 |
| Hospital 3 (including 1 APP pilot) | West | 3 | 1 | 5 | 2 | 11 |
| Hospital 4 | Northeast | 2 | 1 | 1 | 4 | 8 |
| Hospital 5 | South | 1 | 2 | 4 | 0 | 7 |
| Hospital 6 | South | 0 | 3 | 0 | 0 | 3 |
| Hospital 7 (pilot) | West | 0 | 0 | 1 | 0 | 1 |
| Hospital 8 (pilot) | West | 1 | 0 | 0 | 0 | 1 |
| All | National | 17 | 7 | 18 | 7 | 49 |

[a]MD: medical doctor.

[b]APP: advanced practice provider.

[c]RN: registered nurse.

[d]RT: registered therapist.

Of the participating EDs, 4 EDs were in the West (3 Pacific and 1 Mountain), 2 were in the South (1 South Atlantic and 1 East South Central), 1 ED was in the Midwest (West North Central), and 1 ED was in the Northeast (Middle Atlantic). Some hospitals



XSL•FO
RenderX

had 2 focus groups within their EDs, while others only had one. Overall, clinicians averaged 11.2 (SD 7.5) years of experience in their profession as ED clinicians (mean 9.4, SD 7.9 for physicians; mean 11, SD 4.3 for APPs; mean 12, SD 8.1 for RNs; and mean 14, SD 7.5 for RTs; Table 2). In addition, 67% (33/49) identified as female, 71% (35/49) identified as White, and 6% (3/49) identified as Hispanic.

During focus groups, we asked participants about examples of innovations they or their institutions adopted and later rejected. We asked them to describe how the decision was made to reject the use of the innovation at the *individual*, *departmental*, and *institutional* levels. Textbox 2 lists examples of discontinued COVID-19 care innovations mentioned by participants in focus groups, such as intubation boxes, plasma, PPE reuse protocols, and nasal steroids.

**Table .** Participant demographic characteristics.

| Respondent type | Number of respondents | Average years of experience, mean (SD) |
|---|---|---|
| APP[a] | 7 | 11 (4.3) |
| Female | 5 | 11.8 (5.0) |
| Male | 2 | 9 (1.4) |
| ED[b] physician | 17 | 9.4 (7.9) |
| Female | 6 | 6.8 (5.7) |
| Male | 11 | 10.8 (8.8) |
| Nurse | 18 | 12 (8.1) |
| Female | 16 | 11.4 (7.1) |
| Male | 2 | 16.5 (17.7) |
| Respiratory therapist | 7 | 14 (7.5) |
| Female | 6 | 12.8 (7.5) |
| Male | 1 | 21 (N/A)[c] |
| Grand total | 49 | 11.2 (7.5) |

[a]APP: advanced practice provider.

[b]ED: emergency department.

[c]N/A: not available.

**Textbox 2.** Examples of discontinued COVID-19 care innovations.

- Intubation boxes/shields
- Early intubation
- Plasma
- Deep cleaning COVID-19 patient rooms
- Varieties of COVID-19 tests
- V60 (ventilator type)
- PPE reuse protocols
- Nasal steroids and fluids

We present findings related to innovation deimplementation in Table 3, with subsections on factors and processes related to deimplementation. Tables 3 and 4 highlight quotations that are representative of the major themes. We then identified the factors related to the deimplementation of COVID-19 care innovations that participants shared during focus groups.



XSL•FO
RenderX

**Table .** Identifying and prioritizing innovations for deimplementation: themes and illustrative quotations.

| Themes | Illustrative quotation |
|---|---|
| Poor patient outcomes | *By the time we were regularly dealing with severely ill patients, it was evident from the experience—largely out of other [ED[a] group] sites but also somewhat early in the publications—that early intubation actually turned out to be a really bad idea due to bad outcomes. So, we were actually I think early adopters of not [doing] early intubation in relation to our tertiary care receiving facilities.* -MD[b], West |
| Little to no evidence of change | *I think it was a little bit of kind of seeing your patients with the plasma, it never really seemed to change anything. And there was also a lot of data coming out that was showing plasma, a) doesn't really seem to work and even there've been a couple of studies that were showing harm with plasma…I think we kind of just were like, eh, it doesn't really seem to be doing much and the patients it maybe gets some gains or it's very minimal and so let's just stop wasting our time with it.* -MD, Midwest |
| Burdensome | *I honestly don't recall there being a discussion, like, oh, these [intubation boxes] are so cumbersome to use, we're just not going to use them anymore. I don't, I think we all probably just got comfortable with our PPE[c] and I think we were pretty lucky. So maybe there are things that happened that I'm not aware of but it seems like we all did a really good job with PPE.* -MD, Midwest |
| Compliance | *I feel like we just realized nobody is using [telemedicine]. And that's kind of how you know that you kind of stop that process. So I think that's one, is just realizing that nobody is using that equipment or following those processes anymore. In the ED if it works people will usually stick with it or create a workaround as needed. But telemedicine is really, really hard for us to do….* -RN[d], West |
| Patient volume | *I think a lot of [discontinuing innovations] was if we were going to surge or not. So, they would have contingency plans, based on national predictions if things were going to surge or not. So they would be like, okay, we're having lower volumes now, so we're going to start discontinuing this. But in two weeks if it starts going back up, this is the order in which we're going to start things back.* -MD, South |
| Availability | *Yeah, the only thing I stopped doing was double using the N95. Because now it's available, so I don't have to. And another thing we did was to...you know, they provided more supplies which was good, so you don't have to reuse the N95. And you don't have to reuse the surgical mask which is good. And it makes everyone feel more comfortable.* -APP[e], South |
| Leadership | *I think one of our doctors has an ER throughput committee, and so that was brought up in that throughput committee. And he discussed it with infection prevention. And then was able through leadership to make that decision. Like, we [stopped deep cleaning rooms] before but it wasn't official and then people were questioning it. So that's when the committee got involved.* -RN, Midwest |
| New information | *…In the case of intubation boxes, it was a pain to use and then I think there was just like an abstract where they had done an airflow study and they showed how airflow happened and that it actually directed more viral particles back onto the [person performing the procedure] or something like that. And so then that was just the excuse that everyone needed to let us be like 'Oh, thank God,' and we got rid of that thing.* -MD, West |

[a]ED: emergency department.

[b]MD: medical doctor.

[c]PPE: personal protective equipment.

[d]RN: registered nurse.

[e]APP: advanced practice provider.



XSL•FO
RenderX

**Table .** Processes related to the deimplementation of innovations: themes and illustrative quotations.

| Themes | Illustrative quotation |
|---|---|
| Leadership guidance | *I remember that we were really counting on the guidance that we were getting from our leadership in terms of what the next, you know, what the next steps were, and then obviously they were using government resources and what not to pass that guidance on to us.* - APP[a], Northeast |
| Collaboration among colleagues | *And so it all came down to…just [saying], "I think we should hold off" or "hey, I think we should talk to the ICU and see what they think about this because they're the ones who are gonna be managing it later down the road." And then ultimately that dialogue facilitated our own institutional changing in practice and how we approached intubation…it also evolved with the amount of information that we knew about COVID later on, saying that early intubation will not necessarily save someone's life.* -MD[b], West |
| No process/passive | *So I think that's one, is just realizing that nobody is using that equipment or following those processes anymore. Some other things, like we were doing the antibodies for a while, those kind of treatment plans that you don't really ever stop doing but you just realize that you don't have a patient population that fits that requirement anymore. So I can't think of a lot of things that we purposely stop. Things just kind of tend to run their course.* -RN[c], West |

[a]APP: advanced practice provider.

[b]MD: medical doctor.

[c]RN: registered nurse.

## Factors Contributing to Identifying and Prioritizing Innovations for Deimplementation

### *Clinician Experiences*

During focus groups, multiple clinicians noted that the experience of observing poor patient outcomes was a common reason for discontinuing COVID-19 care innovations. For example, participants frequently mentioned early intubation as a deimplemented innovation: early in the pandemic, it was common practice to immediately intubate many patients affected by COVID-19 when they arrived at the ED as this practice was thought to be clinically beneficial. Several focus group participants shared that they stopped early intubation after their direct observation of poor outcomes. As one clinician elaborated:

> *Yeah, so we did try early intubation; we were very aggressive about it. And then I think after we realized that early intubation wasn't necessarily for temporary outcomes in patients who are on the ventilator for a long time and they get complications, learning from our ICU colleagues, and the challenges of the intubation itself are…in post-intubation. it was even harder to deal with, then we scaled it back….* [MD, West]

Witnessing poor outcomes made clinicians reconsider whether it was useful to intervene early with intubation. One participant talked about how the "initial panic" of COVID-19 caused clinicians to initiate intubation early in care processes. The results of this intervention were not always beneficial for the patient. Another clinician highlighted how tracking outcomes and data were important in deciding whether to discontinue an innovation. Relatedly, if a care innovation showed little to no evidence of change in patient outcomes, it was stopped. For example, a participant mentioned that giving patients plasma—an innovation hoped to provide antibodies to sick patients from recovered patients—was an innovation that "never seemed to change anything," with minimal gains. One ED clinician stated that giving plasma never seemed to make a difference for patients:

> *…there was also a lot of data coming out that was showing plasma doesn't really seem to work and even there've been a couple of studies that were showing harm with plasma…I think we kind of just were like, eh, it doesn't really seem to be doing much and the patients it maybe gets some gains or it's very minimal and so let's just stop wasting our time with it.* [MD, Midwest]

Additionally, intubation boxes were frequently brought up as a discontinued innovation because clinicians highlighted that they were cumbersome to use. As new evidence-based information became available, clinicians stopped using them.

### *Contextual Factors and Research Evidence*

Participants reported that once they realized nobody was using the equipment or following procedures for certain innovations, those innovations were eventually discontinued. As one clinician stated, "the innovation[s] just ran [their] course." For example, one clinician reported that telemedicine was an innovation that was never officially discontinued but was eventually stopped due to lack of use. One participant shared that decision-making around discontinuing innovations was based on whether there was a predicted surge or high patient volume. With lower patient volumes, clinicians would discontinue certain innovations, particularly those affecting triage or physical management. Conversely, if they were predicting a high patient volume or a surge, they would restart certain innovations (eg, implementing a surge schedule with overnight coverage).

Relatedly, the availability of the innovation mattered as well. One prominent example was related to reusing PPE. Once there



XSL•FO
RenderX

were adequate supplies of N95 masks available, participants shared that they stopped reusing the PPE:

> *Yeah, the only thing I stopped doing was double using the N95. Because now it's available, so I don't have to. And another thing we did was to...you know, they provided more supplies which was good, so you don't have to reuse the N95. And you don't have to reuse the surgical mask which is good. And it makes everyone feel more comfortable.* [APP, South]

Leadership also played a significant role in identifying and prioritizing discontinued innovations. Across focus groups, participants mentioned that directions regarding innovation deimplementation came from leadership guidance or hospital-wide committees. Oftentimes, official guidance from leadership teams or committees legitimized deimplementation decisions:

> *I think one of our doctors has an ER throughput committee, and so that was brought up in that throughput committee. And he discussed it with infection prevention. And then was able through leadership to make that decision. Like, we [stopped deep cleaning rooms] before but it wasn't official and then people were questioning it. So that's when the committee got involved.* [RN, Midwest]

Finally, as research evidence on different care processes and new innovations emerged, such as high-flow oxygen, clinicians realized that they were able to disregard early intubation and its associated risks and replace it with other modalities. A clinician noted that utilizing other innovations led to the deimplementation of early intubation:

> *…as time went by, we started to use high flow [oxygen], we started other modalities and were able to avoid [early] intubation.* [RT, Northeast]

Early intubation quickly ceased to be viewed as a standard of care and was described as a discontinued innovation by multiple focus groups.

### Processes Related to the Deimplementation of Innovations

#### *Leadership Guidance*

The process of deimplementing innovations occasionally included some sort of leadership or committee. The decisions were made and applied hospital-wide, usually in tandem with Centers for Disease Control and Prevention guidance or other government resources. As one participant shared, "We had a couple APPs who started using [ivermectin] in the beginning. And then once we got some leadership direction, it stopped. But we had a couple that did." Another clinician underscored the importance of leadership for guidance during the process of discontinuing innovations:

> *I remember that we were really counting on the guidance that we were getting from our leadership in terms of what the next steps were…* [APP, Northeast]

#### *Collaboration Among Colleagues*

Collaboration between hospitalists and colleagues was another commonly discussed topic during focus groups when asked about deimplementation. For example, one clinician stated:

> *[with] every COVID patient, we were really thinking out loud. We were just kind of having open conversations about disposition for each of them, just because it wasn't standardized. It wasn't like a kidney stone that we kind of knew the process. I think every patient was truly thought out loud and it was really kind of a team-based decision.* [APP, South]

Additionally, during early intubation, clinicians shared how they would discuss the process with other clinicians in the intensive care unit and determined what worked best for the patient. A clinician described how the approach of early intubation evolved and was eventually discontinued—the process involved multiple factors, such as discussing with colleagues in the intensive care unit, sharing information, and examining outcomes:

> *And so it all came down to…just [saying]… "Hey, I think we should talk to the ICU and see what they think about this because they're the ones who are going to be managing it later down the road." And then ultimately that dialogue facilitated our own institutional changing in practice…it also evolved with the amount of information that we knew about COVID later on, saying that early intubation will not necessarily save someone's life.* [MD, West]

#### *No Process/Passive Deimplementation*

Sometimes, however, there was no process to officially deimplement innovations; it was passive. One participant shared that they stopped using monoclonal antibodies because the patient population for it had declined. The clinician stated,

> *So I think that's one, is just realizing that nobody is using that equipment or following those processes anymore…Some other things, like we were doing the antibodies for a while, those kind of treatment plans that you don't really ever stop doing but you just realize that you don't have a patient population that fits that requirement anymore. So I can't think of a lot of things that we purposely stop.* [RN, West]

In these cases, it seems that enough people simply stop doing something, without any central decision or authority, to the point that it is no longer used.

## Discussion

### Principal Findings

While there have been calls to identify and reduce low-value care before [18,33,34] and during [35] the COVID-19 pandemic, this study adds to what is known about innovation deimplementation during the pandemic by characterizing the factors and processes that led ED clinicians to decide to discontinue using COVID-19 care innovations. Overall, clinicians highlighted that patient outcomes, collaboration and discussion with ED colleagues, supply availability, and protocols



XSL•FO
RenderX

from leadership as factors that contributed heavily to decisions about sustaining or deimplementing innovations in EDs during the COVID-19 pandemic.

Prior studies show that, in routine conditions, enablers of effective low-value care deimplementation include leadership commitment, provider engagement, provider training, performance feedback to providers, and shared decision-making with patients [18,35]. Under usual conditions, deimplementation involves strong external support, provider engagement, and training and education to impact knowledge and behavior; therefore, it is typically a slow process [36]. Fundamental to its goals, the deimplementation of health services involves engaging with and changing the behavior of health care providers. A recent systematic review of determinants of deimplementation found that health care provider knowledge, expectations, attitudes, and behaviors are key determinants of deimplementation [37]. In addition, a central concern of health care providers is the trust that the deimplementation effort is necessary and will not harm patients currently receiving a low-value treatment, while also improving patient health in the long term. This requires building support and buy-in for deimplementation from front-line staff [38]. Clinicians need both motivation and support to properly facilitate deimplementation. Our findings confirm the importance of leadership, collaboration, and discussion with ED colleagues for the deimplementation of newly adopted innovations.

However, the speeds and processes for innovation deimplementation differ significantly during a public health emergency compared to more stable time periods in terms of scientific discoveries related to health. For example, hydroxychloroquine was once thought to be potentially helpful before evidence ultimately discouraged its use. Bradley et al [39] examined hydroxychloroquine deimplementation and found that the proportion of hospitalized patients receiving hydroxychloroquine in 2020 decreased drastically over the span of a few months (55.2% in March and April, 4.8% in May and June, and 0.8% in July and August). The authors hypothesize that this was because hydroxychloroquine was not deeply entrenched in hospital practice; the deimplementation of long-standing innovations may be more challenging due to the sunk cost fallacy and loss aversion [40]. Clinicians may have been more amenable to rapid deimplementation due to awareness of their own limited knowledge, along with the limited evidence-based information and the rapidly changing information during the pandemic, the latter of which was a factor related to innovation deimplementation that we identified during focus groups.

Although there are some conceptual frameworks and criteria for innovation and intervention deimplementation, they may not adequately describe the factors affecting deimplementation during rapidly evolving situations, such as the COVID-19 pandemic [22]. Much of the deimplementation literature focuses on stopping long-standing low-value care, whereas during the pandemic, many discontinued practices were short-term and context-dependent as new evidence emerged [9,10,18,33]. The pace and timing of discontinuing innovations in ED settings during COVID-19 was a major distinguishing factor. The 3 criteria posed by McKay et al [9] to identify if interventions are appropriate for deimplementation (ie, interventions that are "not effective or harmful," "not the most effective or efficient to provide," and "that are no longer necessary") also may fall short in describing conditions during the COVID-19 pandemic because decisions about innovations had to be made in the absence of evidence-based information about innovation efficacy, harm, and efficiency.

Overall, innovation deimplementation in ED settings during the COVID-19 pandemic was different from typical deimplementations due to a lack of information as well as the rapid evolution of information in response to emerging and dynamic environments. Clinicians tended to discontinue the innovation if leadership support, evidence-based information, funding, or resource availability were limited. We also found differences between innovations that were actively stopped due to poor outcomes and new evidence emerging, such as early intubation and intubation boxes, and those that were stopped due to passive processes or a lack of necessity. Prior deimplementation models and frameworks do not account for circumstances where deimplementation is unintentional. Passive deimplementation of innovations may require additional clinician autonomy and greater flexibility, whereas active deimplementation involves highly formalized and structured pathways. Active deimplementation is more difficult compared to passive deimplementation because passive deimplementation occurs without follow-up and drastic changes to practice.

These findings have implications for public health emergency situations. Our focus groups suggest that institutions will be most able to quickly adopt and deimplement innovations informed by local conditions, such as leadership, patient volume, replacement, new information, compliance, and availability, as described above. While prior research identified several other factors that are linked to deimplementation, such as fear of medical malpractice, revenue, competitive advantage, and liability [18], those factors were less relevant in the high-pressure fast-paced environment at the beginning of the pandemic. Our findings also show that factors around deimplementation during a public health emergency were highly localized and akin to a pilot-testing approach—innovations were often hospital-specific and performed at a local scale. Using limited knowledge, clinicians were willing to try and adopt innovations rapidly within their organizations, and the level of evidence needed to implement the innovations was much lower. However, when new evidence-based information became available, clinicians also had to be willing to deimplement those tested innovations quickly. Future public health emergencies will require rapid evaluation and will likely be driven by a narrower group of factors than during normal times.

At the same time, as seen through the Greenhalgh framework, overarching factors that are important for implementation and deimplementation in normal times are still needed during public health emergencies, such as stakeholder engagement, leadership buy-in, and organizational readiness [11,36]. In the future, we should prepare for and embrace the adoption of innovations with limited data, with the willingness to easily deimplement as those data appear and with the knowledge that the deimplementation of innovations used for shorter periods of time may be easier [38,40,41].



XSL•FO
RenderX

### Limitations

Because our focus groups were conducted 2 to 3 years after March 2020, recall bias may have influenced ED participants' self-reported experiences. Additionally, despite recruiting from a diverse set of EDs, our sample is not nationally representative, and we do not have information on those who declined to participate in the focus groups. We also do not believe our results are generalizable, but they have been influential in informing the development of a nationwide survey that will be administered to ED physicians and nurses about implementation and deimplementation during the COVID-19 pandemic. Some of the focus groups were coded by only 1 coder, but the interrater reliability of dually coded transcripts was high [42]. Finally, at the time of writing this article, many COVID-19 care innovations implemented and deimplemented in the early stages of the pandemic now have more evidence to support their benefit or lack thereof.

### Conclusion

ED clinicians are critical to the response to disease outbreaks, such as the COVID-19 pandemic. During uncertain times, clinicians must rapidly implement new protocols and practices and assess these changes in real time under uncertain and potentially dangerous conditions. Stopping unhelpful or harmful practices appropriately is equally important. This study is unique because it examined sustainability and deimplementation during a time of uncertainty and identified unique factors and processes that affect the decision to deimplement innovations, such as poor outcomes, patient volume, availability, and leadership support, when trying to choose the best approach to treating an acute and serious disease.

### Acknowledgments

The authors would like to acknowledge and thank the participating emergency department clinicians who offered their time and insights to the study. Generative artificial intelligence was not used in manuscript preparation.

### Funding

This study was supported by the National Institute of Allergy and Infectious Diseases, within the National Institutes of Health, an agency of the United States Department of Health and Human Services (NIAD5R01AI169293).

### Data Availability

Data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

### Authors' Contributions

Conceptualization: PM, SF, CB
Data curation: CB, SF, CC, NQ, SH
Formal analysis: NQ, SH, CB, SF, CC
Funding acquisition: PM, SF, CB
Investigation: NQ, SH, CB, SF, CC
Methodology: PM, SF, CB
Project administration: NQ, SH, CB, SF, CC
Resources: NQ, SH, CB, SF, CC
Software: NQ, SH
Supervision: PM, SF, CB
Visualization: NQ, SH
Writing – original draft: NQ, SH
Writing – review & editing: SF, PM, CC, CB

### Conflicts of Interest

None declared.

Checklist 1
COREQ checklist.
[PDF File, 203 KB - i-jmr_v15i1e82088_app1.pdf ]

### References

1. Backhouse A, Ogunlayi F. Quality improvement into practice. BMJ 2020 Mar 31;368:m865. [doi: 10.1136/bmj.m865] [Medline: 32234777]



XSL•FO
RenderX

2. Caputo ND, Strayer RJ, Levitan R. Early self-proning in awake, non-intubated patients in the emergency department: a single ED's experience during the COVID-19 pandemic. Acad Emerg Med 2020 May;27(5):375-378. [doi: 10.1111/acem.13994] [Medline: 32320506]
3. Mellado-Artigas R, Ferreyro BL, Angriman F, et al. High-flow nasal oxygen in patients with COVID-19-associated acute respiratory failure. Crit Care 2021 Feb 11;25(1):58. [doi: 10.1186/s13054-021-03469-w] [Medline: 33573680]
4. Bauchner H, Fontanarosa PB, Livingston EH. Conserving supply of personal protective equipment-a call for ideas. JAMA 2020 May 19;323(19):1911. [doi: 10.1001/jama.2020.4770] [Medline: 32196543]
5. Barbash IJ, Sackrowitz RE, Gajic O, et al. Rapidly deploying critical care telemedicine across states and health systems during the Covid-19 pandemic. NEJM Catal Innov Care Deliv 2020;1(4) [FREE Full text]
6. Begley JL, Lavery KE, Nickson CP, Brewster DJ. The aerosol box for intubation in coronavirus disease 2019 patients: an in-situ simulation crossover study. Anaesthesia 2020 Aug;75(8):1014-1021. [doi: 10.1111/anae.15115] [Medline: 32397008]
7. Yaffee AQ, Peacock E, Seitz R, et al. Preparedness, adaptation, and innovation: approach to the COVID-19 pandemic at a decentralized, quaternary care department of emergency medicine. West J Emerg Med 2020 Sep 25;21(6):63-70. [doi: 10.5811/westjem.2020.8.48624] [Medline: 33052812]
8. Prasad V, Ioannidis JP. Evidence-based de-implementation for contradicted, unproven, and aspiring healthcare practices. Implement Sci 2014 Jan 8;9:1. [doi: 10.1186/1748-5908-9-1] [Medline: 24398253]
9. McKay VR, Morshed AB, Brownson RC, Proctor EK, Prusaczyk B. Letting go: conceptualizing intervention de-implementation in public health and social service settings. Am J Community Psychol 2018 Sep;62(1-2):189-202. [doi: 10.1002/ajcp.12258] [Medline: 29971792]
10. Walsh-Bailey C, Tsai E, Tabak RG, et al. A scoping review of de-implementation frameworks and models. Implement Sci 2021 Nov 24;16(1):100. [doi: 10.1186/s13012-021-01173-5] [Medline: 34819122]
11. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629. [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
12. Neumar RW, Blomkalns AL, Cairns CB, et al. Emergency medicine research: 2030 strategic goals. Acad Emerg Med 2022 Feb;29(2):241-251. [doi: 10.1111/acem.14367] [Medline: 34363718]
13. Bernstein SL, Stoney CM, Rothman RE. Dissemination and implementation research in emergency medicine. Acad Emerg Med 2015 Feb;22(2):229-236. [doi: 10.1111/acem.12588] [Medline: 25640627]
14. van Bodegom-Vos L, Davidoff F, Marang-van de Mheen PJ. Implementation and de-implementation: two sides of the same coin? BMJ Qual Saf 2017 Jun;26(6):495-501. [doi: 10.1136/bmjqs-2016-005473] [Medline: 27512102]
15. Helfrich CD, Rose AJ, Hartmann CW, et al. How the dual process model of human cognition can inform efforts to de-implement ineffective and harmful clinical practices: a preliminary model of unlearning and substitution. J Eval Clin Pract 2018 Feb;24(1):198-205. [doi: 10.1111/jep.12855] [Medline: 29314508]
16. Patey AM, Grimshaw JM, Francis JJ. Changing behaviour, "more or less": do implementation and de-implementation interventions include different behaviour change techniques? Implement Sci 2021 Feb 25;16(1):20. [doi: 10.1186/s13012-021-01089-0] [Medline: 33632274]
17. Green T, Bosworth HB, Coronado GD, et al. Factors affecting post-trial sustainment or de-implementation of study interventions: a narrative review. J Gen Intern Med 2024 May;39(6):1029-1036. [doi: 10.1007/s11606-023-08593-7] [Medline: 38216853]
18. Norton WE, Chambers DA. Unpacking the complexities of de-implementing inappropriate health interventions. Implement Sci 2020 Jan 9;15(1):2. [doi: 10.1186/s13012-019-0960-9] [Medline: 31915032]
19. Nilsen P, Ingvarsson S, Hasson H, von Thiele Schwarz U, Augustsson H. Theories, models, and frameworks for de-implementation of low-value care: a scoping review of the literature. Implement Res Pract 2020;1:2633489520953762. [doi: 10.1177/2633489520953762] [Medline: 37089121]
20. Wang V, Maciejewski ML, Helfrich CD, Weiner BJ. Working smarter not harder: coupling implementation to de-implementation. Healthc (Amst) 2018 Jun;6(2):104-107. [doi: 10.1016/j.hjdsi.2017.12.004] [Medline: 29279297]
21. Raudasoja AJ, Falkenbach P, Vernooij RWM, et al. Randomized controlled trials in de-implementation research: a systematic scoping review. Implement Sci 2022 Oct 1;17(1):65. [doi: 10.1186/s13012-022-01238-z] [Medline: 36183140]
22. Niven DJ, Mrklas KJ, Holodinsky JK, et al. Towards understanding the de-adoption of low-value clinical practices: a scoping review. BMC Med 2015 Oct 6;13:255. [doi: 10.1186/s12916-015-0488-z] [Medline: 26444862]
23. Kitzinger J. Qualitative research. Introducing focus groups. BMJ 1995 Jul 29;311(7000):299-302. [doi: 10.1136/bmj.311.7000.299] [Medline: 7633241]
24. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544. [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
25. Geographic levels. United States Census Bureau. 2024. URL: https://www.census.gov/programs-surveys/economic-census/guidance-geographies/levels.html [accessed 2024-06-03]




XSL•FO
RenderX

26. Huilgol SS, Berdahl CT, Qureshi N, Cohen CC, Mendel P, Fischer SH. Innovation adoption, use and implementation in emergency departments during the COVID-19 pandemic. BMJ Innov 2024 Jul;10(3):51-59. [doi: 10.1136/bmjinnov-2023-001206] [Medline: 40785952]
27. Guest G, Namey E, McKenna K. How many focus groups are enough? Building an evidence base for nonprobability sample sizes. Field methods 2017 Feb;29(1):3-22. [doi: 10.1177/1525822X16639015]
28. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. Health Serv Res 2007 Aug;42(4):1758-1772. [doi: 10.1111/j.1475-6773.2006.00684.x] [Medline: 17286625]
29. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960 Apr;20(1):37-46. [doi: 10.1177/001316446002000104]
30. Saldaña J. The Coding Manual for Qualitative Researchers: SAGE Publications; 2015.
31. Dedoose. URL: https://www.dedoose.com/ [accessed 2026-05-06]
32. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
33. Qureshi N, Huilgol SS, Berdahl CT, Cohen CC, Mendel P, Fischer SH. Understanding uptake of information about innovations among emergency department clinicians during the COVID-19 pandemic. J Am Coll Emerg Physicians Open 2024 Aug;5(4):e13243. [doi: 10.1002/emp2.13243] [Medline: 39035811]
34. Norton WE, Kennedy AE, Chambers DA. Studying de-implementation in health: an analysis of funded research grants. Implement Sci 2017 Dec 4;12(1):144. [doi: 10.1186/s13012-017-0655-z] [Medline: 29202782]
35. McKay VR, Kwon JH. Identifying, deconstructing, and deimplementing low-value infection control and prevention interventions. Infect Control Hosp Epidemiol 2023 Jun;44(6):1034-1035. [doi: 10.1017/ice.2023.53] [Medline: 37183993]
36. Gangathimmaiah V, Drever N, Evans R, et al. What works for and what hinders deimplementation of low-value care in emergency medicine practice? A scoping review. BMJ Open 2023 Nov 9;13(11):e072762. [doi: 10.1136/bmjopen-2023-072762] [Medline: 37945299]
37. Augustsson H, Ingvarsson S, Nilsen P, et al. Determinants for the use and de-implementation of low-value care in health care: a scoping review. Implement Sci Commun 2021 Feb 4;2(1):13. [doi: 10.1186/s43058-021-00110-3] [Medline: 33541443]
38. Shelton RC, Brotzman LE, Johnson D, Erwin D. Trust and mistrust in shaping adaptation and de-implementation in the context of changing screening guidelines. Ethn Dis 2021;31(1):119-132. [doi: 10.18865/ed.31.1.119] [Medline: 33519162]
39. Bradley SM, Emmons-Bell S, Mutharasan RK, et al. Repeated cross-sectional analysis of hydroxychloroquine deimplementation in the AHA COVID-19 CVD Registry. Sci Rep 2021 Jul 23;11(1):15097. [doi: 10.1038/s41598-021-94203-7] [Medline: 34302004]
40. Kulkarni SA, Leykum LK, Moriates C. Deimplementation: discontinuing low-value, potentially harmful hospital care. J Hosp Med 2021 Jan;16(1):63. [doi: 10.12788/jhm.3563] [Medline: 33357335]
41. Watanabe JH, Kwon J, Nan B, Abeles SR, Jia S, Mehta SR. Medication use patterns in hospitalized patients with COVID-19 in California during the pandemic. JAMA Netw Open 2021 May 3;4(5):e2110775. [doi: 10.1001/jamanetworkopen.2021.10775] [Medline: 34019090]
42. Cole R. Inter-rater reliability methods in qualitative case study research. Sociol Methods Res 2024 Nov;53(4):1944-1975. [doi: 10.1177/00491241231156971]


## Abbreviations

**APP:** advanced practice provider
**COREQ:** Consolidated Criteria for Reporting Qualitative Research
**ED:** emergency department
**MD:** medical doctor
**PPE:** personal protective equipment
**RN:** registered nurse
**RT:** respiratory therapist



XSL•FO
RenderX

*Edited by M Balcarras; submitted 08.Aug.2025; peer-reviewed by R Guo, S Raza; revised version received 18.Mar.2026; accepted 09.Apr.2026; published 22.May.2026.*

*Please cite as:*
*Huilgol S, Qureshi N, Berdahl C, Cohen C, Mendel P, Fischer S*
*Innovation Deimplementation in Emergency Departments During the COVID-19 Pandemic: Qualitative Study of Clinicians' Decision-Making*
*Interact J Med Res 2026;15:e82088*
*URL: https://www.i-jmr.org/2026/1/e82088*
*doi:10.2196/82088*

© Shreya Huilgol, Nabeel Qureshi, Carl Berdahl, Catherine Cohen, Peter Mendel, Shira Fischer. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 22.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Sociodemographic and Health-Risk Determinants of COVID-19 Vaccine Booster Preferences and Willingness to Pay in Singapore: Discrete Choice Experiment

Hooi Swang Cheng, MSc; Yi Wang, PhD; Sharon Hui Xuan Tan, MPH; Ian Yi Han Ang, PhD

Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Block MD1, 12 Science Drive 2, Tahir Foundation Building, #09-01, Singapore, Singapore

**Corresponding Author:**
Hooi Swang Cheng, MSc
Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Block MD1, 12 Science Drive 2, Tahir Foundation Building, #09-01, Singapore, Singapore

## Abstract

**Background:** Waning immunity and recurring COVID-19 waves point to the importance of sustained booster vaccination strategies. With repeated boosters needed to sustain population-level immunity, evaluating public preferences for COVID-19 booster vaccination is important for informing future vaccination strategies, as changing perceptions of urgency, threat, and motivation may alter willingness to receive vaccine boosters compared with initial uptake.

**Objective:** This study aimed to explore the public's stated preferences and implicit willingness to pay for COVID-19 vaccine boosters in Singapore and to identify the sociodemographic factors and health-risk acceptance profiles associated with these outcomes.

**Methods:** A discrete choice experiment survey was conducted among the general population in Singapore, recruited from an online panel in June 2022. Discrete choice experiment scenarios presented hypothetical scenarios varying by booster characteristics, including effectiveness, protection duration, and cost, and risks of COVID-19 infection, hospitalization, and death. A 2-stage discrete choice design was used: participants first selected between 2 hypothetical scenarios and then reported real-life booster acceptance of their chosen option. A mixed logit model was used for analyses. We also estimated implicit willingness to pay for vaccine boosters, as well as assessed the associations of sociodemographic factors and health-risk profiles with booster acceptance or hesitancy.

**Results:** A total of 1567 participants were included in the analysis. Overall, respondents preferred boosters that offered a longer duration of protection, particularly under higher risks of COVID-19 infection and death. Respondents were willing to pay S$1.43 for every percentage-point increase in booster effectiveness, S$4.95 (S$1=US $0.77 as of July 10, 2026) per additional month of protection, and an additional S$55.23 for a vaccine booster under the hypothetical worst-case scenario of COVID-19. Younger adults, women, individuals of non-Chinese ethnicity, those without children, people with higher socioeconomic status, individuals who had received 2 or fewer COVID-19 vaccine doses, and those with chronic conditions or on chronic disease medication were more likely to receive a booster. On the other hand, older adults, individuals with lower education or socioeconomic status, those with prior COVID-19 infection, and those exhibiting higher health-risk acceptance were less likely to accept booster vaccination.

**Conclusions:** Our findings demonstrate how vaccine attributes, sociodemographic factors, and health-risk attitudes influence booster acceptance. These insights can guide policymakers in tailoring strategies and communication efforts that address population needs and behavioral drivers to improve booster uptake, address hesitancy, and support sustainable long-term immunization coverage, contributing to adaptive vaccination planning in the ongoing management of COVID-19.

***(Interact J Med Res 2026;15:e87909)*** doi:10.2196/87909

**KEYWORDS**

population preferences; booster vaccination; social determinants; vaccine acceptance; vaccine hesitancy

## Introduction

Over 3 years into the COVID-19 pandemic, the World Health Organization declared the end of the Public Health Emergency of International Concern on May 5, 2023 [1]. As the pandemic transitioned into an endemic phase, many countries resumed prepandemic norms. However, the shift to endemicity does not imply that SARS-CoV-2 has become less transmissible or less virulent. Historical experiences with past influenza pandemics have shown that periodic outbreaks can occur during the endemic phase, influenced by factors such as the emergence of new viral variants, changes in population immunity, vaccine



XSL•FO
RenderX

coverage, and the implementation of public health and social measures [2].

Epidemiological studies show that COVID-19 exhibits a cyclical pattern of waves during its endemic phase [2,3]. Furthermore, according to surveillance data on SARS-CoV-2 activity globally monitored by the Global Influenza Surveillance and Response System, there have been periodic peaks observed over the past few years [4]. In May 2025, several Asian regions, including Hong Kong, Thailand, Singapore, and China, reported a resurgence of COVID-19 cases [4-6]. Health authorities attributed this surge to the emergence of a highly transmissible COVID-19 variant, increased social interactions during festivities, and waning population immunity [6,7]. These developments underscore the ongoing threat posed by SARS-CoV-2 and highlight the need for sustained immunological protection.

Protective immunity through vaccination remains critical in the fight against COVID-19. However, evidence shows that vaccine-induced immunity wanes over time, potentially leading to increased susceptibility to infection and severe outcomes [8,9]. This has reinforced the importance of routine COVID-19 booster vaccinations to maintain sufficient population-level immunity and mitigate severe outcomes, particularly among vulnerable groups. Therefore, it is imperative to better understand the public's attitudes toward COVID-19 vaccine boosters.

COVID-19 vaccines have demonstrated varying levels of effectiveness and durability against infection, hospitalization, and death [10]. Understanding the trade-offs associated with COVID-19 vaccine boosters—such as the balance between vaccine effectiveness, durability, risks associated with COVID-19, and booster out-of-pocket costs—is critical for informing adaptive and effective vaccination strategies. These insights can guide public health decisions on optimal booster intervals, target populations, and allocation of health care resources, while also accounting for opportunity cost and public acceptability. In this study, we assessed these trade-offs using a discrete choice experiment (DCE), which is a preference elicitation technique commonly used to determine the relative importance of different factors influencing people's choices when choosing between alternatives. Several DCE studies have explored public preferences for COVID-19 vaccinations [11], but there is a lack of studies focusing on booster vaccination decisions [12,13]. Public acceptance of boosters may shift from initial vaccination uptake as perceived urgency and threat decline, pandemic fatigue sets in, and trust in vaccine safety evolves [14]. Reassessing public preferences during the transition from pandemic response to endemic management is important for anticipating future booster uptake under changing conditions. Such evidence is crucial for informing public health policies, optimizing booster schedules, and tailoring communication strategies to sustain population-level immunity.

This study aimed to explore the public's stated preferences and implicit willingness to pay (WTP) for COVID-19 vaccine boosters in Singapore and to identify the sociodemographic factors and health-risk acceptance profiles associated with these outcomes. The results will enable differentiated outreach and health-risk communication based on various profiles of vaccine booster acceptance. They will also allow policymakers, community leaders, and researchers to identify key target groups and roll out targeted measures and interventions to increase vaccine booster uptake.

## Methods

### Study Participants

The study was a cross-sectional online survey administered through the Singapore Population Health Studies Online Panel, approved by the National University of Singapore Institutional Review Board (NUS-IRB; reference number: H-18 - 011). This panel served as an online research platform that facilitated public health research on a diverse range of topics through monthly surveys related to public health. When the survey was conducted, the panel comprised 2058 active participants who were demographically representative of the general population of Singapore.

### Setting

Singapore is a city-state nation located in Southeast Asia with a population of over 5.7 million. Following the first COVID-19 case detected in January 2020, Singapore increased social distancing measures, with an eventual 2-month national partial lockdown (termed a circuit breaker) in April 2020 [15]. The COVID-19 vaccines were first introduced in Singapore in December 2020 [16]. Subsequently, a COVID-19 vaccine booster program was introduced in Singapore in September 2021 for older adults, residents of aged care facilities, and individuals with weakened immune systems, which was then gradually expanded to all other age groups [17]. In January 2022, it was announced that all individuals aged 18 and older would need to receive a vaccine booster no later than 270 days after their primary vaccination series to maintain their fully vaccinated status [18]. The survey was administered in June 2022, during the period when this policy was still in force.

As the pandemic situation stabilized, Singapore announced the easing of COVID-19 measures and the transition to an endemic phase of COVID-19 in February 2023 [19]. The nation has since been living with the endemic new normal; however, routine COVID-19 vaccination was still regarded as an important key preventive measure against severe COVID-19 illness and hospitalization [19,20].

### Survey Structure

The survey was designed with 2 main parts. The first part presented participants with two sets of DCEs, each with seven choice tasks, to elicit (1) the public's level of acceptance of and WTP for COVID-19 vaccine boosters and (2) the nonpharmaceutical intervention measures they would be inclined to trade off if they did not want to receive and/or pay for the vaccine boosters. This manuscript focuses on the discussion of the first DCE, as the second DCE is of a different study design, with the results published elsewhere [21].

The second part of the survey consisted of 41 questions that collected sociodemographic information and assessed the level of risk acceptance and tolerance in the health domain using the



XSL•FO
RenderX

Health Risk Attitude Scale (HRAS-13). The HRAS-13 is a psychometric instrument measuring risk attitudes in the health domain [22], which has demonstrated strong reliability and has been used in various population surveys and WTP studies [23,24].

## DCE Design

### *Selection of Attributes and Levels*

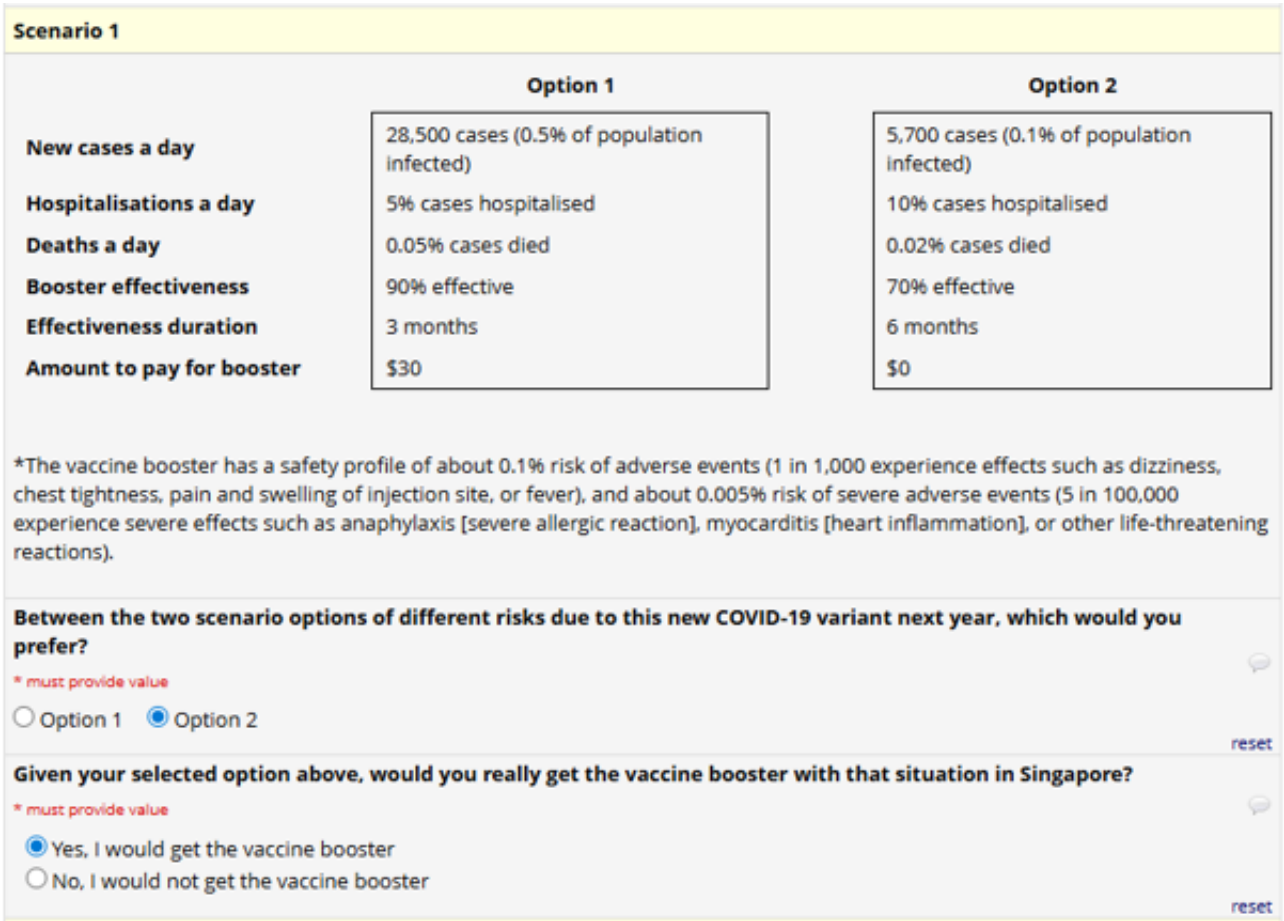
The DCE scenarios presented a combination of different levels of the various factors of interest, focusing on vaccine booster characteristics against different risks of COVID-19 in Singapore, with the following attributes tested: (1) new cases a day: personal risk of COVID-19 infection, (2) hospitalizations a day: personal risk of hospitalization due to COVID-19, (3) deaths a day: personal risk of death from COVID-19, (4) booster effectiveness: effectiveness of the booster shot against the dominant strain, (5) effectiveness duration: duration of the booster shot maintained its level of efficacy, and (6) amount to pay for booster: the out-of-pocket booster cost.

The selection of attributes corresponding to the risks of COVID-19 was based on commonly reported statistics in news articles and press releases by the health authorities (ie, daily numbers of new COVID-19 cases, hospitalizations, and deaths), with which the respondents would be familiar [25]. The levels for the COVID-19 risk attributes were determined based on reasonable event rates relative to Singapore's population size, supplemented by informal consultations with researchers familiar with the local epidemiological context. The attribute levels for vaccine booster's effectiveness and effectiveness duration were selected based on what was similarly presented in the literature [26,27].

### *Experimental Design*

The DCE questions were designed using Sawtooth Software Inc's Lighthouse Studio (version 9.8) [28]. The DCE used an unlabeled choice-based conjoint design with a "dual-response none" approach. Participants were presented with a choice task consisting of 2 scenarios for them to choose in the first stage, followed by a second-stage question regarding a "no-choice option." In the second stage, participants were asked whether they would get themselves vaccinated with a booster shot in real life given the COVID-19 situation illustrated in the option they selected in the first stage. Vaccine booster acceptance and hesitancy were defined based on participants' responses to the second-stage question of the DCE tasks. Participants who answered "Yes, I would get the vaccine booster" were classified as demonstrating booster acceptance, while those who answered "No, I would not get the vaccine booster" were classified as demonstrating booster hesitancy. An example of a DCE choice task is presented in Figure 1.

**Figure 1.** Sample of discrete choice experiment choice task.




A simulation exercise was conducted to examine the coverage matrix of the design and test the required sample size. A preliminary power analysis was conducted for the 2 DCEs. A rule of thumb has been proposed for determining the minimum



XSL•FO
RenderX

sample size for aggregate-level full-profile choice-based conjoint modeling: n ≥500 *c*/*ta*, where n is the number of respondents, *t* is the number of choice tasks, *a* is the number of alternatives per task (not including the none alternative), and *c* is the largest number of levels for any attribute [29,30]. By this formula, the required number of respondents was 167. We tested the DCE designs using the Test Design function within the software. The projected standard errors for each attribute level of both DCEs were all less than 0.05.

The DCE survey underwent 2 rounds of user acceptability testing to obtain feedback on suitability, ease of understanding and phrasing, and survey length. The survey and the DCEs were refined before being rolled out to the full panel.

Each participant was randomly assigned to one of the 20 DCE questionnaire sets. Each questionnaire set had 6 unique choice tasks, totaling 120 unique choice tasks, and an additional dominant choice task for a logic check. As this was an online self-administered survey and the DCE questions can be relatively complicated compared with typical survey questions, we took a precautionary step to ensure the quality of responses by applying a dominance test, whereby one of the two choice alternatives was clearly superior. The dominance task helped to identify participants who did not understand the DCE questions or responded randomly without meaningful processing of the choice tasks. Participants who failed the dominance test were excluded from the analysis.

## Statistical Analysis

The DCE data were analyzed using a mixed logit model [31] in the following 3 steps.

### *Overall Preferences for COVID-19 Vaccine Boosters and WTP*

First, a mixed logit model was used to understand participants' overall level of acceptance of and WTP for vaccine boosters, as well as their preferences for COVID-19 vaccine boosters based on booster effectiveness and effectiveness duration. The mixed logit model considers preference heterogeneity for the attributes of the alternatives while accounting for random preference heterogeneity at the individual level.

The random utility $U_{ij}$ of individual *i*, for attribute *j*, where $x_{ij}$ is a vector of observed variables describing the attributes of the alternatives $_{ij}$ and is the idiosyncratic error term assumed to follow a type 1 extreme value distribution, was specified as

(1)Uij=βi×xij+ ij

where the term is

(2)βi=β+ηi

The $\eta_i$ term follows a multivariate normal distribution representing the random preference heterogeneity, which includes all unobserved factors that affect individuals' preferences. The model assumes a linear combination of each utility component, including additional variables for (1) "none option," which captures the participants' responses in the second stage of each choice task and (2) left-and-right bias, which accounts for the tendency of participants to select only options presented on the left or the right [32].

The COVID-19 risk attributes were treated as discrete variables in the model. The price and characteristics of the vaccine booster were treated as continuous variables in the model based on the tabulation pattern of the levels of each attribute and the percentage of selections, which demonstrated a linear trend. Model selection was based on the Bayesian information criterion, comparing the model using continuous values with the model using discrete values for the 3 attributes of price, effectiveness, and protection duration of the vaccine booster, one variable at a time. The model with the minimum Bayesian information criterion value was selected.

The WTP was estimated as the ratio of the utility parameter for each attribute to the cost attribute, which was the out-of-pocket payment for a vaccine booster under the hypothetical context of this study. These estimates represent the marginal WTP for a change in a specific attribute level. For example, a WTP estimate for vaccine effectiveness reflects the additional amount respondents were willing to pay for an increase in effectiveness relative to the reference level. The WTP was presented in Singapore dollars (S$) denomination.

(3)WTP=−βjβbooster price

### *Preferences for COVID-19 Vaccine Boosters When Facing Higher Risks of COVID-19*

Second, the mixed logit model included 2-way interactions between the characteristics of the vaccine booster and the risks associated with COVID-19 to examine how preferences for the vaccine booster's characteristics varied with the risks of COVID-19 infection, hospitalization, and death. The 2 attributes, booster effectiveness and booster effectiveness duration, were interacted with the highest levels of the risks of COVID-19 attributes (ie, 57,000 new cases [1% of the population infected], 10% cases hospitalized, and 0.1% of cases died per day, respectively). The preference weight coefficients of the interaction terms indicated the changes in preference weights as the levels of booster effectiveness and booster effectiveness duration increased under each scenario of COVID-19 situation.

### *Demographic Factors Associated With COVID-19 Vaccine Booster Acceptance or Hesitancy and Characteristics of Vaccine Booster*

Third, the mixed logit model was used to assess the impact of demographic factors on the level of acceptance of and WTP for vaccine boosters, and on preferences for vaccine boosters based on their characteristics. Hence, the $\beta_i$ term in Equation (1) was specified as

(4)βi=β+γzi+ηi

where is the demographic variables of individual *i* that influence the mean of each attribute coefficient, and γ is the matrix of parameters measuring differential preferences by demographic variables. The impact of demographic factors on each attribute was considered one at a time. Only the terms with *P* values <.10 were kept in the final model. In the final model, positive coefficients indicate lower utility for booster acceptance and, therefore, greater hesitancy to accept the booster relative to the reference group, while negative coefficients indicate



XSL•FO
RenderX

higher utility for booster acceptance and, therefore, greater willingness to accept the booster relative to the reference group.

Demographic factors that were considered included gender, ethnicity, age group, marital status, education level, employment status, housing type, family with children, family with children younger than 5 years, COVID-19 vaccination status, personal history of COVID-19 infection, having chronic diseases and taking medications for chronic diseases, and HRAS-13 score measuring health-related risk attitudes. Age was classified into 3 categories: younger than 40 years, 40 to 60 years, and older than 60 years. Ethnicity was classified into 2 categories: Chinese and non-Chinese (Malay, Indian, or others). Marital status was classified into 2 categories: married and not married (single, divorced, or widowed). Participants with a secondary school education or below were considered to have lower education, while participants with postsecondary school or tertiary education levels were considered to have higher education. Employment status was classified into 2 categories: full-time and part-time or not working. Information on housing type was used as a proxy for socioeconomic status (SES) [33], which was classified into 3 categories: 1- to 3-room public housing (low SES), 4- to 5-room public housing (middle SES), and private housing (high SES).

The HRAS-13 score was derived from a set of 13 questions, with participants rating their attitudes toward health risks on a 7-point Likert scale. The score was calculated by summing the 13 questions (some were reverse scored), and a higher score is indicative of having a higher level of risk acceptance in the health-related domain. The HRAS-13 score was modeled as a continuous variable.

For the analysis of booster acceptance or hesitancy, the base reference population consisted of those who were male, of Chinese ethnicity, aged 40 to 60 years, married with children but had no children under the age of 5 years, full-time workers, of middle SES, with a lower education level, had no previous COVID-19 infection, had received at least 3 shots of COVID-19 vaccine, had no chronic diseases, and had an HRAS-13 score of 0.

The level of statistical significance was set at $P<.05$. All statistical analyses were carried out using R software (version 4.2.2) [34].

### Ethical Considerations

Singapore Population Health Studies Online Panel members had previously provided informed consent as part of a master ethical approval to participate in the panel (NUS-IRB reference number: H-18 - 011). The members received notification by email and/or SMS text messages about this survey on June 20, 2022, to access the survey via the provided link and were given 10 days to complete the survey on the REDCap platform. At the start of the online survey, participants were presented with a participant information sheet. Participation in the survey was voluntary and anonymous, and participants only proceeded to the survey questions after acknowledging that they had read and received a copy of the participant information sheet. Ethical approval and a waiver of documented informed consent were obtained for this study from the National University of Singapore Institutional Review Board (reference number: NUS-IRB-2020 - 82). To ensure anonymity, no personal identifiers were linked to the survey responses. Participants were reimbursed S$15 (S$1=US $0.77 as of July 10, 2026) upon survey completion.

## Results

### Participant Characteristics

A total of 1651 survey responses were received from the 2058 active participants on the online research platform, corresponding to a response rate of 80.2%. Of the 1651 survey respondents, 84 were excluded; 83 respondents failed the dominance test, and 1 respondent indicated an age less than 18 years old. Among the 1567 included respondents (Table 1), the majority were female (n=914, 58.3%), of Chinese ethnicity (n=1337, 85.3%), married (n=999, 63.8%), had no chronic disease (n=1238, 79%), had attained postsecondary or tertiary education (n=1211, 77.3%), lived in 4- to 5-room public housing (n=1005, 64.1%), had children (n=917, 58.5%), and worked full-time (n=877, 56.0%). Most of them reported that they had received at least 3 doses of COVID-19 vaccine (n=1443, 92.1%) and had not contracted COVID-19 (n=940, 60%). Slightly less than half of the included respondents (n=700, 44.7%) were aged 40 to 60 years. The HRAS-13 score had a mean of 41.89 (SD 7.90); a higher score indicates higher health-risk acceptance.

Table 2 shows the appearance and selection frequencies for the different levels of the attributes used in the DCE. In about one-fifth of the choice tasks, respondents chose not to accept the vaccine booster given their options made in the first stage of the choice task.



XSL•FO
RenderX

**Table .** Demographic characteristics of the included respondents (N=1567).

| Demographic factors | Participants |
|---|---|
| Gender, n (%) | |
| Female | 914 (58.3) |
| Male | 653 (41.7) |
| Ethnicity, n (%) | |
| Chinese | 1337 (85.3) |
| Non-Chinese (Malay, Indian, and others) | 230 (14.7) |
| Age groups (y), n (%) | |
| <40 | 499 (31.8) |
| 40 - 60 | 700 (44.7) |
| >60 | 368 (23.5) |
| Marital status, n (%) | |
| Married | 999 (63.8) |
| Not married (single, divorced, separated, or widowed) | 530 (33.8) |
| Prefer not to say | 38 (2.4) |
| Have children, n (%) | |
| Yes | 917 (58.5) |
| No | 616 (39.3) |
| Prefer not to say | 34 (2.2) |
| Have children younger than 5 y, n (%) | |
| Yes | 146 (9.3) |
| No | 1387 (88.5) |
| Prefer not to say | 34 (2.2) |
| Highest level of education | |
| Lower education (secondary or below) | 324 (20.7) |
| Higher education (postsecondary or tertiary) | 1211 (77.3) |
| Prefer not to say | 32 (2.0) |
| Employment status, n (%) | |
| Full-time | 877 (56.0) |
| Part-time or not working | 690 (44.0) |
| Housing type, n (%) | |
| 1- to 3-room public housing (low SES[a]) | 337 (21.5) |
| 4- to 5-room public housing (middle SES) | 1005 (64.1) |
| Private housing (high SES) | 179 (11.4) |
| Others/prefer not to say/do not know | 46 (2.9) |
| COVID-19 vaccination status, n (%) | |
| Received 2 shots and below | 124 (7.9) |
| Received 3 shots or above | 1443 (92.1) |
| Previous COVID-19 infection, n (%) | |
| Yes | 619 (39.5) |
| No | 948 (60.5) |
| Having a chronic disease, n (%) | |
| Yes | 322 (20.5) |



XSL•FO
RenderX

| Demographic factors | Participants |
|---|---|
| No | 1245 (79.5) |
| On chronic medications, n (%) | |
| Yes | 269 (17.2) |
| No | 1298 (82.8) |
| Health Risk Attitude Scale-13 score, mean (SD) | 41.89 (7.90) |

[a]SES: socioeconomic status.

**Table .** The various attributes and their levels in the discrete choice experiment, along with the appearance and selection frequencies of the different levels for the included respondents.

| Attributes and levels | Appearances, N | Selection, n (%) |
|---|---|---|
| Daily new cases | | |
| 5700 cases (0.1% of the population infected) | 6263 | 3554 (56.7) |
| 28,500 cases (0.5% of the population infected) | 6189 | 2998 (48.4) |
| 57,000 cases (1% of the population infected) | 6352 | 2850 (44.9) |
| Daily hospitalizations | | |
| 1% of cases hospitalized | 6272 | 3590 (57.2) |
| 5% of cases hospitalized | 6272 | 3100 (49.4) |
| 10% of cases hospitalized | 6260 | 2712 (43.3) |
| Daily deaths | | |
| 0.02% of cases died | 6263 | 3543 (56.6) |
| 0.05% of cases died | 6272 | 3228 (51.5) |
| 0.10% of cases died | 6269 | 2631 (42.0) |
| Booster effectiveness | | |
| 50% effective | 6274 | 1812 (28.9) |
| 70% effective | 6264 | 3193 (51.0) |
| 90% effective | 6266 | 4397 (70.2) |
| Booster effectiveness duration (mo) | | |
| 3 | 6270 | 2403 (38.3) |
| 6 | 6287 | 3459 (55.0) |
| 9 | 6247 | 3540 (56.7) |
| Booster price (S$) | | |
| 0 | 4714 | 2982 (63.3) |
| 10 | 4694 | 2627 (56.0) |
| 30 | 4693 | 2160 (46.0) |
| 50 | 4703 | 1633 (34.7) |
| None option (unwilling to take vaccine booster) | 9402 | 2045 (21.8) |

## Overall Preferences for COVID-19 Vaccine Boosters and WTP

Table 3 summarizes the utility, or preference weight, from the mixed logit model. There was no evidence of left-and-right bias in choice selection (left=0.02, 95% CI –0.05 to 0.09; $P$=.48). Overall, the level of acceptance of the vaccine booster was high (coefficient for the none option=–6.58, 95% CI –6.91 to –6.24; $P$<.001), with a predicted uptake probability of 99.9% (95% CI 99.8% to 99.9%; $P$<.001). However, the random preference heterogeneity at the individual level for vaccine booster acceptance or hesitancy was large (SD for the none option=5.63, 95% CI 5.24 to 6.02; $P$<.001), indicating heterogeneous preferences among the respondents with respect to their decisions regarding accepting the vaccine booster in the real-world decision-making contexts.



XSL•FO
RenderX

The pattern of the estimated utilities for the attribute levels of risks of COVID-19 and vaccine booster characteristics was consistent with the expected utility theory [35]. Higher disutilities were exhibited for higher levels of risks of COVID-19, with increasing numbers of daily new cases, hospitalizations, and deaths. There were stronger preferences for a vaccine booster that offered stronger protection against COVID-19 in terms of its effectiveness and the duration of protection from COVID-19 variants. The relative utility increased with booster effectiveness (coefficient=0.05, 95% CI 0.04 to 0.05; *P*<.001) and booster effectiveness duration (coefficient=0.16, 95% CI 0.14 to 0.18; *P*<.001).

The WTP was generated using the coefficient of booster price as the base. The average WTP value for every percentage-point increase in booster effectiveness was S$1.43, and S$4.95 for every month increase in the duration of booster effectiveness. Respondents were willing to pay an additional S$55.23 for a vaccine booster under the hypothetical worst-case scenario of COVID-19 as presented in the DCE.

**Table .** Mixed logit model regression results.

| Variable | Coefficient (95% CI) | *P* value | WTP[a] |
|---|---|---|---|
| Left | 0.02 (−0.04 to 0.09) | .48 | —[b] |
| None option | −6.58 (−6.91 to −6.24) | <.001 | — |
| 57,000 new cases | −0.60 (−0.69 to −0.51) | <.001 | −18.59 |
| 28,500 new cases | −0.30 (−0.39 to −0.21) | <.001 | −9.32 |
| 10% of cases hospitalized | −0.53 (−0.62 to −0.43) | <.001 | −16.20 |
| 5% of cases hospitalized | −0.26 (−0.35 to −0.17) | <.001 | −7.96 |
| 0.10% of cases died | −0.66 (−0.76 to −0.57) | <.001 | −20.45 |
| 0.05% of cases died | −0.26 (−0.35 to −0.18) | <.001 | −8.15 |
| Booster effectiveness | 0.05 (0.04 to 0.05) | <.001 | 1.43 |
| Booster effectiveness duration | 0.16 (0.14 to 0.18) | <.001 | 4.95 |
| Booster price | −0.03 (−0.03 to −0.03) | <.001 | — |

[a]WTP: willingness to pay.

[b]Not applicable.

## Preferences for COVID-19 Vaccine Booster When Facing Higher Risks of COVID-19

Results shown in Table 4 indicate that under the scenarios with the highest numbers of daily new cases and daily deaths, individuals' utility for booster effectiveness duration further increased (coefficient=0.06, *P*<.005, respectively). However, people became more price sensitive in the scenario with the highest number of daily deaths (coefficient=−0.01, 95% CI −0.01 to −0.00; *P*=.04).



XSL•FO
RenderX

**Table .** Mixed logit model with 2-way interactions between characteristics of vaccine booster and the highest levels of COVID-19 risk as illustrated in the discrete choice experiment.

| Variable | Coefficient | *P* value |
|---|---|---|
| Left | 0.04 (−0.03 to 0.11) | .25 |
| None option | −6.55 (−6.91 to −6.20) | <.001 |
| 57,000 new cases | −0.77 (−1.32 to −0.22) | .006 |
| 28,500 new cases | −0.35 (−0.45 to −0.26) | <.001 |
| 10% of cases hospitalized | −0.23 (−0.77 to 0.32) | .42 |
| 5% of cases hospitalized | −0.27 (−0.36 to −0.17) | <.001 |
| 0.10% of cases died | −1.06 (−1.58 to −0.54) | <.001 |
| 0.05% of cases died | −0.30 (−0.39 to −0.20) | <.001 |
| Booster effectiveness | 0.05 (0.04 to 0.05) | <.001 |
| Booster effectiveness duration | 0.13 (0.10 to 0.15) | <.001 |
| Booster price | −0.03 (−0.04 to −0.03) | <.001 |
| 57,000 new cases × booster effectiveness | −0.00 (−0.01 to 0.00) | .28 |
| 57,000 new cases × booster effectiveness duration | 0.06 (0.02 to 0.10) | .004 |
| 57,000 new cases × booster price | 0.00 (−0.00 to 0.01) | .47 |
| 10% of cases hospitalized × booster effectiveness | −0.00 (−0.01 to 0.00) | .23 |
| 10% of cases hospitalized × booster effectiveness duration | −0.01 (−0.05 to 0.04) | .75 |
| 10% of cases hospitalized × booster price | 0.00 (−0.00 to 0.01) | .79 |
| 0.10% of cases died × booster effectiveness | 0.00 (−0.00 to 0.01) | .48 |
| 0.10% of cases died × booster effectiveness duration | 0.06 (0.02 to 0.09) | .003 |
| 0.10% of cases died × booster price | −0.01 (−0.01 to −0.00) | .04 |

## Demographic Factors Associated With COVID-19 Vaccine Booster Acceptance or Hesitancy and Characteristics of Vaccine Booster

Of the 1567 included respondents, 1469 provided complete demographic information and were included in this analysis. Ninety-eight respondents were excluded due to “Prefer not to say” responses. Demographic factors associated with vaccine booster acceptance were female, non-Chinese ethnicity, aged younger than 40 years, having no children, high SES (defined as living in private housing), having received 2 shots of COVID-19 vaccine or less, having a chronic disease, and on chronic disease medication. On the other hand, vaccine booster hesitancy was noticed among those aged older than 60 years, with a lower education level, low SES (defined as living in 1- to 3-room public housing), who had previously contracted COVID-19, and with higher health-risk acceptance (see Figure 2). However, there were no strong associations observed between demographics and characteristics of vaccine booster. Results are provided in Multimedia Appendix 1.



XSL•FO
RenderX

**Figure 2.** Association between demographic factors and COVID-19 vaccine booster acceptance or hesitancy. HRAS: health-risk attitude scale score; SES: socioeconomic status. ***$P$<.001; *$P$<.05.

## Discussion

### Principal Results

This study examined the associations between sociodemographic characteristics, individual health-risk profiles, and public preferences for COVID-19 vaccine boosters, including out-of-pocket cost, among the Singapore population using a DCE. Overall, our study showed that booster preferences varied across sociodemographic and health-risk subgroups. Individuals with lower educational attainment (coefficient=2.22, 95% CI 1.83-2.63), older adults (coefficient=0.59, 95% CI 0.34-0.85), and those living in smaller public housing types (coefficient=0.34, 95% CI 0.01-0.69) showed lower acceptance of vaccine boosters, underscoring persistent socioeconomic gradients in booster uptake. Health-risk profiles also played an important role; we found that individuals with higher health-risk acceptance were less likely to accept booster vaccination (coefficient=2.37, 95% CI 1.77-2.97). These findings provide important insights into the key drivers of booster vaccination decisions and highlight critical subgroups that may require tailored communication and policy interventions to sustain booster uptake.

### Comparisons With Prior Work

Respondents were willing to pay more to gain the benefit, that is, an additional S$4.95 for every month increase in the duration of vaccine booster effectiveness. These findings imply that the risks of COVID-19 may influence people's WTP for vaccines. Respondents were willing to pay a much higher amount (ie, an additional S$55.23) for vaccine boosters to avoid the hypothetical worst-case scenario of COVID-19. The estimated average WTP value for every unit percent increase in booster effectiveness was S$1.43, which was much lower than that reported in the Australian study for COVID-19 vaccine (ie, Aus $23.92) [36]. This finding was not unexpected, as people are less willing to pay for the vaccine out of pocket considering the recurring costs of repeated vaccine boosters required. Additionally, Singapore has provided and continues to provide the COVID-19 vaccines and the booster doses free of charge.

In the multivariable analysis, we found that the health-risk profile and several sociodemographic characteristics, including age, gender, ethnicity, education level, SES, having children, chronic conditions, COVID-19 vaccination status, and previous history of COVID-19 infection, were associated with vaccine booster acceptance or hesitancy. Our findings identified that individuals with higher health-risk acceptance were less likely to accept a vaccine booster and preferred a vaccine booster with lower effectiveness. Moreover, our findings of greater booster acceptance among those who did not complete their primary vaccination series, had a chronic disease, and were taking medication for a chronic disease also provide evidence that individuals with higher risks of COVID-19 infection had a higher utility for vaccine boosters.

We found that individuals who reported that they had contracted COVID-19 were more hesitant to accept the vaccine booster. A study showed that natural immunity provides a comparable



XSL•FO
RenderX

level of protection against new COVID-19 infection as vaccine-induced immunity [37]. Those with a previous history of COVID-19 infection might feel confident in the protection by infection-induced immunity and/or not be too concerned about the effects of an infection. As such, they might deem an additional booster dose as redundant [38]. Interestingly, we found that individuals with no children were more likely to accept boosters, suggesting that individuals with and without young dependents may perceive the risks and benefits of vaccine boosters differently for adults and children. As of April 2023, the Singapore Ministry of Health reported that more than two-thirds of children aged 5 to 11 years had not completed the vaccine doses required for minimum protection against COVID-19 [39]. Despite having confidence and trust in the COVID-19 vaccine for adults, parents, who are often the decision makers of child vaccination, may have concerns about the safety of the COVID-19 vaccine for their children and may reject booster doses for themselves when they are hesitant to vaccinate their children [40,41].

Our findings corroborate previous literature showing greater booster hesitancy among those with lower educational attainment [41-43] and greater booster acceptance among those with high SES groups [42,44] and those with a chronic disease [45]. On the other hand, we found some contrary results compared to previous studies. Our findings identified that there was greater booster acceptance among females, while the opposite was previously found in a meta-analysis [42]. A previous study conducted in Singapore found no gender differences in COVID-19 vaccine hesitancy [46]. However, greater vaccine acceptance was observed among females in a population-based study conducted in Hong Kong [47].

Although most studies have reported lower COVID-19 booster hesitancy among older adults [41,42,48], this study found that booster hesitancy was higher among older adults, whereas younger individuals exhibited greater acceptance of COVID-19 booster doses in the multivariable-adjusted model. Our findings are consistent with a recent systematic review and a study from Malaysia that found higher levels of booster hesitancy among older adults, suggesting that age-related differences in booster acceptance may vary across populations and contexts [43,49]. Furthermore, recent national data from Singapore indicated that, as of June 2024, booster vaccination rates among older adults aged 60 years and older remained low, lending further support to our findings [50]. Greater effort should be put forth on educational campaigns and in targeting public health messaging to individuals of older age, lower educational attainment, and lower SES to improve vaccine booster uptake among these subpopulations. This study also found that individuals of non-Chinese ethnicity were more likely to accept vaccine boosters, suggesting that the efforts made by the government and Muslim religious authorities to reduce vaccination barriers among minority groups have been effective in addressing earlier hesitancy [46].

The predicted booster uptake probability observed in this study was high, nearly 100%. Respondents may have been influenced by the prevailing policy environment, specifically the booster vaccination recommendations and the requirement to maintain fully vaccinated status, both of which were introduced in early 2022 and remained in effect at the time of data collection. This may partly reflect the broader policy environment, as well as social desirability bias in responses, whereby respondents reported choices that aligned with perceived public health expectations rather than reflecting their actual future behavior. It may also be attributable to the characteristics of the study sample, which was highly vaccinated: 92% of respondents self-reported having received three or more vaccine doses at the time of the survey. This suggests that respondents were already predisposed toward vaccine acceptance, and the near-universal predicted uptake may therefore reflect a pre-existing provaccination orientation and prior vaccination behavior, rather than the effects of the DCE attributes alone.

As with all DCE studies, our estimates reflect stated rather than revealed preferences. A positive attitude toward a vaccine booster does not necessarily always translate to actual uptake in reality [45]. The near-universal predicted uptake observed in this study contrasts sharply with actual booster uptake in Singapore, where about half among seniors aged 60 years and older had received an additional booster dose by November 2023 [51], and fewer than 20% of the local population had received a COVID-19 booster dose within the preceding year by May 2024 [52]. This discrepancy may also reflect the influential role of policy mandates and perceived disease risk in shaping vaccination behavior. During the initial booster rollout, uptake was high, with approximately 84.5% of eligible individuals having received a first booster dose by January 2022 [53], likely reflecting the combined effects of vaccination-differentiated measures and heightened concern about COVID-19. When these measures were withdrawn and the perceived threat of COVID-19 declined, uptake of subsequent annual boosters fell markedly.

On the other hand, the discrepancy between stated preferences and actual booster uptake may partly reflect factors beyond vaccine attributes and sociodemographic characteristics. Previous studies have highlighted the influence of psychological determinants, including safety, vaccine confidence or side effects, and trust, which may affect whether favorable intentions ultimately translate into vaccination behavior [54,55].

## Limitations

This study has several limitations. First, as this was a cross-sectional study, it captured a snapshot of public sentiment at a single point in time and could not account for potential changes in attitudes over time. As most countries have now transitioned to the endemic phase of COVID-19, public attitudes toward vaccine boosters are likely to evolve. Given that the study was conducted before the formal transition to the endemic phase, the findings may not be fully generalizable to the current endemic context. However, this timing allowed the study to capture public preferences during a critical phase when booster policy was being adapted from pandemic response to longer-term management under evolving risk perceptions and policy signals. Future research employing longitudinal designs would be valuable for monitoring shifts in public perceptions and acceptance of boosters in response to changing epidemiological conditions or policy cycles.



XSL•FO
RenderX

One other limitation is that DCEs may be subject to hypothetical bias. Nevertheless, we sought to reduce hypothetical bias by emphasizing the application to real-life choices. Future studies can assess the concordance between stated and revealed preferences within individuals [56], and expand on cheap-talk scripts to further reduce hypothetical bias [57]. Another limitation is that the worst-case scenarios included in the DCE (57,000 daily new cases, 10% of cases hospitalized, and 0.10% of cases died) exceeded the reported daily case burden observed in Singapore and may have been perceived as less realistic by some respondents. However, these levels were designed to represent a severe outbreak scenario and to capture preferences across a broad range of epidemiological conditions. Nevertheless, respondents who failed the dominant choice task were excluded from the analysis, providing some reassurance that participants remained engaged with and were able to meaningfully evaluate the hypothetical scenarios.

Finally, although the participant pool was demographically representative of the Singaporean population, the study was limited to those who were able to read and understand English. About 20% of the resident population aged 15 years and older are illiterate in English, with over half of them being older adults aged 55 years and older [58]. Future research could include other languages to capture individuals not included in the current study and improve the generalizability of the findings. This could be especially important to further inform comprehensive policy decision-making aimed at more targeted public health messaging. Despite these limitations, the study offers valuable insights into the determinants of booster acceptance and provides evidence to inform vaccination planning as COVID-19 management continues to evolve.

## Conclusions

Vaccine boosters are a vital tool to maintain societal resilience against COVID-19. Our findings demonstrated that individuals generally preferred boosters offering longer-lasting protection, particularly in the context of heightened COVID-19 risks of infection and death. These findings underscore the dynamic nature of public attitudes, which may shift in response to evolving policy measures, vaccination recommendations, and epidemiological trends. In addition, this study identified the factors associated with booster acceptance and hesitancy, providing valuable insights that can shape targeted public health strategies and messaging to promote the uptake of vaccine boosters. These findings may extend beyond COVID-19, offering guidance for the development of vaccination policies for future pandemics and epidemics. The estimates of the WTP values in this study may also guide pricing decisions for COVID-19 vaccination in the long term. As preferences for preventive health interventions are highly context dependent, the results should be interpreted in the context of the period in which the study was conducted and may not fully represent the current setting or true endemic-phase perceptions.

## Acknowledgments

We would like to thank Shumian Yeo and the rest of the Singapore Population Health Studies Online Panel team for supporting the setup, testing, and execution of the survey in this study. We would also like to thank Jason Chin Huat Yap, Qian Yang, Xian Yi Ng, and Jacinta I-Pei Chen for their input and assistance with the user acceptability testing. We would also like to thank Hwee Lin Wee for her advice on this study.

The authors declare the use of generative AI in the research and writing process. According to the GAIDeT taxonomy (2025), the following task was delegated to generative AI tools under full human supervision: proofreading and editing. The generative AI tools used were ChatGPT-4o and GPT-5.5. Responsibility for the final manuscript lies entirely with the authors. Generative AI tools are not listed as authors and do not bear responsibility for the final outcomes. Declaration submitted by HSC.

## Funding

This research was supported by the Ministry of Education, Singapore, under the Academic Research Fund Tier 1 (FY2021) and the SSHSPH Seed Fund for Ground-Up Research 2021. This study was conducted using the Singapore Population Health Studies Online Panel, which was funded through the Singapore Ministry of Health's National Medical Research Council Centre Grant Programme (grant number NMRC/CG/C026/2017_NUHS) for the Singapore Population Health Improvement Centre (SPHERiC).

## Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request, subject to approval by the National University of Singapore Institutional Review Board.

## Authors' Contributions

Conceptualization: IYHA (lead), SHXT (equal), YW (equal)
Data curation: HSC
Formal analysis: HSC
Funding acquisition: IYHA (lead), SHXT (equal)
Investigation: IYHA (lead), SHXT (equal)
Methodology: IYHA (lead), SHXT (equal), YW (equal), HSC (supporting)
Project administration: IYHA



XSL•FO
RenderX

Resources: YW
Software: HSC
Supervision: IYHA
Validation: IYHA
Visualization: HSC
Writing – original draft: HSC
Writing – review & editing: HSC (lead), IYHA (equal), SHXT (supporting), YW (supporting)
All authors have read and approved the manuscript. All authors attest they meet the ICMJE criteria for authorship.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Associations between demographic factors and COVID-19 vaccine booster acceptance/hesitancy and characteristics of vaccine booster using mixed logit model.
[DOCX File, 17 KB - i-jmr_v15i1e87909_app1.docx ]

## References


1. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/europe/emergencies/situations/covid-19 [accessed 2025-05-30]
2. Cohen LE, Spiro DJ, Viboud C. Projecting the SARS-CoV-2 transition from pandemicity to endemicity: epidemiological and immunological considerations. PLoS Pathog 2022 Jun;18(6):e1010591. [doi: 10.1371/journal.ppat.1010591] [Medline: 35771775]
3. Townsend JP, Hassler HB, Lamb AD, et al. Seasonality of endemic COVID-19. mBio 2023 Dec 19;14(6):e0142623. [doi: 10.1128/mbio.01426-23] [Medline: 37937979]
4. COVID-19 - Global situation. World Health Organization. 2025. URL: https://www.who.int/emergencies/disease-outbreak-news/item/2025-DON572 [accessed 2025-06-01]
5. Hong Kong's COVID-19 activity shows signs of decline from its peak but is expected to remain high in near term. The Government of the Hong Kong Special Administrative Region – Press Releases. 2025. URL: https://www.info.gov.hk/gia/general/202505/29/P2025052900785.htm [accessed 2025-06-01]
6. Update on COVID-19 situation. Ministry of Health (Singapore). 2025. URL: https://www.moh.gov.sg/newsroom/update-on-covid-19-situation [accessed 2025-05-30]
7. New Covid-19 wave grips Hong Kong, Singapore, China, Thailand. The Nation. 2025. URL: https://www.nationthailand.com/health-wellness/40050177 [accessed 2025-06-01]
8. Dadras O, SeyedAlinaghi S, Karimi A, et al. COVID-19 vaccines' protection over time and the need for booster doses; a systematic review. Arch Acad Emerg Med 2022;10(1):e53. [doi: 10.22037/aaem.v10i1.1582] [Medline: 36033989]
9. Pooley N, Abdool Karim SS, Combadière B, et al. Durability of vaccine-induced and natural immunity against COVID-19: a narrative review. Infect Dis Ther 2023 Feb;12(2):367-387. [doi: 10.1007/s40121-022-00753-2] [Medline: 36622633]
10. Yap C, Ali A, Prabhakar A, et al. Comprehensive literature review on COVID-19 vaccines and role of SARS-CoV-2 variants in the pandemic. Ther Adv Vaccines Immunother 2021;9:25151355211059791. [doi: 10.1177/25151355211059791] [Medline: 34870090]
11. Hinzpeter EL, Kairies-Schwarz N, Beaudart C, Douxfils J, Nayak D, Hiligsmann M. A systematic review of discrete choice experiments on preferences for COVID-19 vaccinations. Patient 2025 Sep;18(5):461-480. [doi: 10.1007/s40271-025-00753-7] [Medline: 40638063]
12. Raman S, Kriner D, Ziebarth N, Simon K, Kreps S. COVID-19 booster uptake among US adults: assessing the impact of vaccine attributes, incentives, and context in a choice-based experiment. Soc Sci Med 2022 Oct;310:115277. [doi: 10.1016/j.socscimed.2022.115277] [Medline: 36001917]
13. Tran BX, Do AL, Boyer L, et al. Preference and willingness to pay for the regular COVID-19 booster shot in the Vietnamese population: theory-driven discrete choice experiment. JMIR Public Health Surveill 2023 Jan 31;9:e43055. [doi: 10.2196/43055] [Medline: 36599156]
14. Parsons Leigh J, FitzGerald EA, Moss SJ, et al. The evolution of vaccine hesitancy through the COVID-19 pandemic: a semi-structured interview study on booster and bivalent doses. Hum Vaccin Immunother 2024 Dec 31;20(1):2316417. [doi: 10.1080/21645515.2024.2316417] [Medline: 38390696]
15. Circuit breaker to minimise further spread of COVID-19. Ministry of Health (Singapore). 2020. URL: https://www.moh.gov.sg/newsroom/circuit-breaker-to-minimise-further-spread-of-covid-19/ [accessed 2025-06-01]
16. Government accepts recommendations of expert committee on COVID-19 vaccination. Ministry of Health (Singapore). 2020. URL: https://www.moh.gov.sg/newsroom/government-accepts-recommendations-of-expert-committee-on-covid-19-vaccination/ [accessed 2025-06-01]




XSL•FO
RenderX

17. Updates on Expanded Vaccination Programme, Facilitating Our Vaccination Booster Programme. Ministry of Health (Singapore). 2021. URL: https://www.moh.gov.sg/newsroom/updates-on-expanded-vaccination-programme-facilitating-our-vaccination-booster-programme/ [accessed 2025-06-01]
18. Focusing our resources and refining our healthcare protocols to deal with the omicron wave. Ministry of Health (Singapore). 2022. URL: https://www.moh.gov.sg/newsroom/focusing-our-resources-and-refining-our-healthcare-protocols-to-deal-with-the-omicron-wave/ [accessed 2025-06-01]
19. Singapore to exit acute phase of pandemic. Ministry of Health (Singapore). 2023. URL: https://www.moh.gov.sg/newsroom/singapore-to-exit-acute-phase-of-pandemic/ [accessed 2025-06-01]
20. COVID-19 national vaccination programme. Ministry of Health (Singapore). 2025. URL: https://www.moh.gov.sg/seeking-healthcare/overview-of-diseases/covid-19-national-vaccination-programme/ [accessed 2025-05-30]
21. Wang Y, Har CE, Tan SHX, Cheng HS, Ang IYH. Preferences for nonpharmaceutical interventions during the endemic phase of COVID-19: discrete choice experiment. JMIR Public Health Surveill 2025 Jun 4;11:e67725. [doi: 10.2196/67725] [Medline: 40470547]
22. Huls SPI, van Osch SMC, Brouwer WBF, van Exel J, Stiggelbout AM. Psychometric evaluation of the Health-Risk Attitude Scale (HRAS-13): assessing the reliability, dimensionality and validity in the general population and a patient population. Psychol Health 2022 Jan;37(1):34-50. [doi: 10.1080/08870446.2020.1851689] [Medline: 33259250]
23. Himmler S, van Exel J, Perry-Duxbury M, Brouwer W. Willingness to pay for an early warning system for infectious diseases. Eur J Health Econ 2020 Jul;21(5):763-773. [doi: 10.1007/s10198-020-01171-2] [Medline: 32180067]
24. Bansback N, Harrison M, Sadatsafavi M, Stiggelbout A, Whitehurst DGT. Attitude to health risk in the Canadian population: a cross-sectional survey. CMAJ Open 2016;4(2):E284-E291. [doi: 10.9778/cmajo.20150071] [Medline: 27398375]
25. Lim J. 19,420 new covid-19 cases in Singapore, 1,355 patients hospitalised. The Straits Times. 2022. URL: https://www.straitstimes.com/singapore/health/19420-new-cases-in-singapore-1355-covid-19-patients-hospitalised [accessed 2025-07-25]
26. Zeng B, Gao L, Zhou Q, Yu K, Sun F. Effectiveness of COVID-19 vaccines against SARS-CoV-2 variants of concern: a systematic review and meta-analysis. BMC Med 2022 May 23;20(1):200. [doi: 10.1186/s12916-022-02397-y] [Medline: 35606843]
27. Feikin DR, Higdon MM, Abu-Raddad LJ, et al. Duration of effectiveness of vaccines against SARS-CoV-2 infection and COVID-19 disease: results of a systematic review and meta-regression. Lancet 2022 Mar 5;399(10328):924-944. [doi: 10.1016/S0140-6736(22)00152-0] [Medline: 35202601]
28. Lighthouse studio. Sawtooth Software. URL: https://sawtoothsoftware.com/lighthouse-studio [accessed 2025-05-31]
29. Orme B. Sample size issues for conjoint analysis studies. : Sawtooth Software, Inc; 1998 URL: https://sawtoothsoftware.com/resources/technical-papers/sample-size-issues-for-conjoint-analysis-studies [accessed 2026-07-08]
30. Johnson R, Orme B. Getting the most from CBC. : Sawtooth Software, Inc; 2003 URL: https://sawtoothsoftware.com/resources/technical-papers/getting-the-most-from-cbc [accessed 2026-07-08]
31. Sarrias M, Daziano R. Multinomial logit models with continuous and discrete individual heterogeneity in R: the gmnl package. J Stat Softw 2017;79(2):1-46. [doi: 10.18637/jss.v079.i02]
32. Veldwijk J, Marceta SM, Swait JD, Lipman SA, de Bekker-Grob EW. Taking the shortcut: simplifying heuristics in discrete choice experiments. Patient 2023 Jul;16(4):301-315. [doi: 10.1007/s40271-023-00625-y] [Medline: 37129803]
33. Malhotra R, Malhotra C, Chan A, Østbye T. Life-course socioeconomic status and obesity among older Singaporean Chinese men and women. J Gerontol B Psychol Sci Soc Sci 2013 Jan;68(1):117-127. [doi: 10.1093/geronb/gbs102] [Medline: 23161348]
34. R Core Team. R Foundation for Statistical Computing. 2022. URL: https://www.R-project.org [accessed 2026-07-08]
35. Cohen BJ. Is expected utility theory normative for medical decision making? Med Decis Making 1996;16(1):1-6. [doi: 10.1177/0272989X9601600101] [Medline: 8717589]
36. Borriello A, Master D, Pellegrini A, Rose JM. Preferences for a COVID-19 vaccine in Australia. Vaccine 2021 Jan 15;39(3):473-479. [doi: 10.1016/j.vaccine.2020.12.032] [Medline: 33358265]
37. Franchi M, Pellegrini G, Cereda D, et al. Natural and vaccine-induced immunity are equivalent for the protection against SARS-CoV-2 infection. J Infect Public Health 2023 Aug;16(8):1137-1141. [doi: 10.1016/j.jiph.2023.05.018] [Medline: 37267680]
38. Cascini F, Pantovic A, Al-Ajlouni Y, Failla G, Ricciardi W. Attitudes, acceptance and hesitancy among the general population worldwide to receive the COVID-19 vaccines and their contributing factors: a systematic review. EClinicalMedicine 2021 Oct;40:101113. [doi: 10.1016/j.eclinm.2021.101113] [Medline: 34490416]
39. Less than a third of children aged 5 to 11 in Singapore have minimum protection against covid-19: MOH. CNA (Channel NewsAsia). 2023. URL: https://www.channelnewsasia.com/singapore/less-third-children-aged-5-11-in-singapore-have-minimum-protection-against-covid-19-moh-5587931 [accessed 2024-10-29]
40. Byrne A, Thompson LA, Filipp SL, Ryan K. COVID-19 vaccine perceptions and hesitancy amongst parents of school-aged children during the pediatric vaccine rollout. Vaccine 2022 Nov 2;40(46):6680-6687. [doi: 10.1016/j.vaccine.2022.09.090] [Medline: 36220714]
41. Lazarus JV, Wyka K, White TM, et al. A survey of COVID-19 vaccine acceptance across 23 countries in 2022. Nat Med 2023 Feb;29(2):366-375. [doi: 10.1038/s41591-022-02185-4] [Medline: 36624316]




XSL•FO
RenderX

42. Galanis P, Vraka I, Katsiroumpa A, et al. First COVID-19 booster dose in the general population: a systematic review and meta-analysis of willingness and its predictors. Vaccines (Basel) 2022 Jul 8;10(7):1097. [doi: 10.3390/vaccines10071097] [Medline: 35891260]
43. De Araújo JST, Delpino FM, Andrade-Gonçalves RLP, et al. Determinants of COVID-19 vaccine acceptance and hesitancy: a systematic review. Vaccines (Basel) 2024 Nov 29;12(12):1352. [doi: 10.3390/vaccines12121352] [Medline: 39772014]
44. Wee LE, Yap AJW, Dickens B, et al. Access to COVID-19 vaccination by socio-economic status in older Singaporean adults: a population-based cohort study. Public Health 2024 Aug;233:38-44. [doi: 10.1016/j.puhe.2024.05.003] [Medline: 38850601]
45. Abdelmoneim SA, Sallam M, Hafez DM, et al. COVID-19 vaccine booster dose acceptance: systematic review and meta-analysis. Trop Med Infect Dis 2022 Oct 13;7(10):298. [doi: 10.3390/tropicalmed7100298] [Medline: 36288039]
46. Subramaniam M, Abdin E, Shafie S, et al. Sociodemographic and psychosocial factors associated with vaccine hesitancy - results from a longitudinal study in Singapore. Hum Vaccin Immunother 2023 Aug 1;19(2):2235964. [doi: 10.1080/21645515.2023.2235964] [Medline: 37489273]
47. Luk TT, Zhao S, Wu Y, Wong JYH, Wang MP, Lam TH. Prevalence and determinants of SARS-CoV-2 vaccine hesitancy in Hong Kong: a population-based survey. Vaccine 2021 Jun 16;39(27):3602-3607. [doi: 10.1016/j.vaccine.2021.05.036] [Medline: 34034950]
48. Antonini M, Genie MG, Attwell K, et al. Are we ready for the next pandemic? Public preferences and trade-offs between vaccine characteristics and societal restrictions across 21 countries. Soc Sci Med 2025 Feb;366:117687. [doi: 10.1016/j.socscimed.2025.117687] [Medline: 39939032]
49. Wong LP, Alias H, Siaw YL, et al. Intention to receive a COVID-19 vaccine booster dose and associated factors in Malaysia. Hum Vaccin Immunother 2022 Nov 30;18(5):2078634. [doi: 10.1080/21645515.2022.2078634] [Medline: 35648441]
50. Status of COVID-19 vaccination among Singaporeans and reasons for not taking additional booster shots. Ministry of Health (Singapore). 2024. URL: https://www.moh.gov.sg/newsroom/status-of-covid-19-vaccination-among-singaporeans-and-reasons-for-not-taking-additional-booster-shots/ [accessed 2025-05-30]
51. Fourth COVID-19 vaccination dose. Ministry of Health (Singapore). 2023. URL: https://www.moh.gov.sg/newsroom/fourth-covid-19-vaccination-dose/ [accessed 2026-06-05]
52. Update on COVID-19 situation. Ministry of Health (Singapore). 2024. URL: https://www.moh.gov.sg/news-highlights/details/update-on-covid-19-situation [accessed 2024-10-08]
53. Yap AJW, Ponce LJ, Wee LE, et al. Effectiveness of mobile vaccination teams to increase COVID-19 booster uptake in Singapore: a population-based synthetic control study. BMC Glob Public Health 2026 Jan 2;4(1):3. [doi: 10.1186/s44263-025-00233-y] [Medline: 41485032]
54. Roy DN, Biswas M, Islam E, Azam MS. Potential factors influencing COVID-19 vaccine acceptance and hesitancy: a systematic review. PLoS One 2022;17(3):e0265496. [doi: 10.1371/journal.pone.0265496] [Medline: 35320309]
55. Roy DN, Islam E, Hossen MM, Ferdiousi N, Azam MS. Sociopsychological determinants of COVID-19 vaccine acceptance and hesitancy among the students' of higher secondary schools in rural Bangladesh: a cross-sectional study. Psychol Sch 2024 Feb;61(2):568-581. [doi: 10.1002/pits.23069]
56. Quaife M, Terris-Prestholt F, Di Tanna GL, Vickerman P. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. Eur J Health Econ 2018 Nov;19(8):1053-1066. [doi: 10.1007/s10198-018-0954-6] [Medline: 29380229]
57. Smeele NVR, van Cranenburgh S, Donkers B, de Bekker-Grob EW. Do we say what we do? Testing cheap talk to reduce hypothetical bias in moral choice experiments. Soc Sci Med 2026 Jul;400:119305. [doi: 10.1016/j.socscimed.2026.119305] [Medline: 42030730]
58. Singapore Department of Statistics (SINGSTAT). Resident population aged 15 years and over by language literate in, age group, sex and ethnic group (census of population 2020). data.gov.sg. 2024. URL: https://data.gov.sg/datasets/d_6ce74ee5822089b435d3159c54d8caf7/view [accessed 2026-06-18]


## ABBREVIATIONS

**DCE:** discrete choice experiment
**HRAS-13:** Health Risk Attitude Scale-13
**NUS-IRB:** National University of Singapore Institutional Review Board
**SES:** socioeconomic status
**WTP:** willingness to pay



XSL•FO
RenderX

*Edited by M Balcarras; submitted 16.Nov.2025; peer-reviewed by E Islam, T Ricks; revised version received 22.Jun.2026; accepted 23.Jun.2026; published 29.Jul.2026.*

*Please cite as:*
*Cheng HS, Wang Y, Tan SHX, Ang IYH*
*Sociodemographic and Health-Risk Determinants of COVID-19 Vaccine Booster Preferences and Willingness to Pay in Singapore: Discrete Choice Experiment*
*Interact J Med Res 2026;15:e87909*
*URL: https://www.i-jmr.org/2026/1/e87909*
*doi:10.2196/87909*

© Hooi Swang Cheng, Yi Wang, Sharon Hui Xuan Tan, Ian Yi Han Ang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 29.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Health Care–Related Determinants of First-Time Long-Term Care Need in Older Adults in Germany: Retrospective Cohort Study Using Claims Data

Till Baldenius[1,2], MSc; Kathrin Jürchott[1], PhD; Christine Haeger[2], PhD; Susann Behrendt[1], MA; Stefan Blüher[2], PhD; Susanne Schnitzer[2], PhD

[1]AOK Research Institute (WIdO), Rosenthaler Straße 31, Berlin, Germany

[2]Institute of Medical Sociology and Rehabilitation Science, Charité – University Medical Center Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

**Corresponding Author:**
Till Baldenius, MSc
AOK Research Institute (WIdO), Rosenthaler Straße 31, Berlin, Germany

## Abstract

**Background:** In recent years, German long-term care (LTC) insurance has experienced an unprecedented increase in the number of beneficiaries. This raises the question of the role of health care in preventing or delaying the need for LTC. At present, related findings are limited. Addressing this research gap could promote longevity and improve quality of life while reducing the financial strain on the social security system.

**Objective:** This study aimed to investigate the associations between the utilization of health care services and first-time LTC need.

**Methods:** This retrospective cohort study examined nationwide linked claims data from the German statutory health and LTC insurance fund, AOK. The dataset included all individuals aged ≥60 years. Using multiple logistic regression, we investigated the association between health care utilization during the 5-year exposure period from 2016 to 2020 and the occurrence of first-time LTC need during the first quarter of 2021. Physician care, pharmaceutical care, physiotherapy, and medical aids were analyzed while adjusting for age, sex, regional variables, and comorbidities. Metrically scaled variables were categorized using the Fisher-Jenks algorithm to explore possible nonlinearities. Individuals who needed LTC prior to 2021 were excluded.

**Results:** The study population comprised 5.3 million individuals. A total of 54.3% (n=2.9 million) were women, and the mean age was 71.3 (SD 8.16) years. Receiving more than 2 of the 5 recommended screenings and vaccinations examined, compared with receiving none, was associated with a strong reduction in the odds of first-time LTC need (odds ratio [OR] 0.62, 95% CI 0.60 - 0.64; *P*<.001). Odds of first-time LTC need were also significantly lower with high numbers of specialist groups and low numbers of specialist days (more than 7 specialist groups consulted and fewer than 48 billing days) compared with no specialist utilization (any specialist utilization: OR 0.82, 95% CI 0.79 - 0.85; *P*<.001; more than 7 instead of fewer than 5 specialist groups: OR 0.94, 95% CI 0.92 - 0.96; *P*<.001). Likewise, a physiotherapy prescription in between 5 and 10 quarters of the 5-year period instead of none was related to lower odds of first-time LTC need (OR 0.81, 95% CI 0.79 - 0.83; *P*<.001). Generalist care, hospitalizations, polypharmacy, and potentially inadequate medication were concomitant with first-time LTC need. Among the disease management programs and medical aids examined, there were both positive and negative relationships with first-time LTC need.

**Conclusions:** Utilization of recommended screenings, vaccinations, specialist physician care, and physiotherapy is substantially and significantly associated with the nonoccurrence of first-time LTC need in the older German population. Our study provides a foundation for future research on orienting health care toward preventing functional decline.

***(Interact J Med Res 2026;15:e86572)*** doi:10.2196/86572

**KEYWORDS**

patient care; health services for the aged; long-term care; preventive health services; cohort studies; retrospective studies; routinely collected health data; aged; early detection of cancer; vaccination; specialization; ambulatory care; physical therapy modalities; drug prescriptions; Germany; epidemiology



XSL•FO
RenderX

## Introduction

The growing number of people in need of long-term care (LTC) in aging societies is accompanied by significant challenges for health care systems worldwide. As populations age, the prevalence of chronic diseases, functional impairments, and frailty increases, leading to a steadily rising demand for continuous support and health care services [1]. In Germany, the number of people in need of LTC doubled between 2015 and 2023 to reach 5.7 million. The 2.2% quarterly increase during this period was significantly higher than that projected based on demographic trends [2]. For the purposes of this study, the need for LTC is defined as a condition in which an individual requires significant assistance with activities of daily living and, as a result, successfully applies for benefits under Germany's statutory LTC insurance scheme [3].

The increasing number of people in need of LTC can be partly attributed to catch-up effects in LTC assessments following the COVID-19 pandemic. Additionally, the so-called "sandwich effect" is under discussion: a simultaneous rise in care-dependent baby boomers and their very old parents. This phenomenon is driven by higher life expectancy and improved survival of individuals with severe illnesses or injuries who require LTC services [4]. This trend is observed not only in Germany but also across industrialized countries worldwide, especially those where population aging is already well advanced, as a recent UN World Social Report states [5]. Although not all older individuals require formal care or support, the overall societal demand for LTC services is rising significantly. In Germany, it is projected that the number of people in need of LTC will rise from 6.0% in 2021 to 9.1% in 2055 [6,7].

Given this trend, the promotion of health and prevention in older age is becoming increasingly important. The goal is to maintain functional health—and thus autonomy and quality of life—for as long as possible, ultimately preventing or delaying the onset of LTC need. Prevention in this context encompasses a wide range of strategies, including health promotion, early detection, and rehabilitative measures. Existing research has largely focused on behavioral preventive approaches aimed at improving physical or cognitive performance among older populations [8]. In contrast, there has been relatively little investigation into structural or health care–related determinants of LTC need [9].

To our knowledge, the role of health care utilization in the years leading up to the onset of LTC need remains an unexplored area, apart from 2 observational studies from Japan, which reported that hyperpolypharmacy, defined as 10 or more medications per month, and prescription of drugs with sedative or anticholinergic properties were associated with an elevated risk of first-time LTC need [10,11]. An ecological Japanese study found that decentralized local prevention services, such as functional training, health education, and support for social activities, significantly reduced the population risk of moderate functional dependency [12]. These findings are specific to the Japanese context and either focus solely on medication or use data at the regional level. The lack of German studies analyzing the association of intersectoral health care utilization with LTC need warrants further individual-level investigation.

This study addresses this research gap by aiming to investigate the associations between the utilization of health care services and first-time LTC need. We analyzed this relationship in individuals aged ≥60 years using linked German statutory health insurance (SHI) and statutory LTC insurance claims data.

## Methods

### Data and Study Design

This retrospective cohort study used anonymized claims data of individuals insured with Germany's largest SHI and statutory LTC insurance fund, AOK, from 2016 to 2021. In 2024, AOK provided health insurance to one-third of the German population, and 42% of all those in need of LTC were insured with AOK [13]. To qualify as in need of LTC, individuals first must formally apply for LTC benefits. The subsequent LTC need assessment considers both mental and physical limitations to independent living. If 1 of 5 care levels is awarded after the assessment, the person becomes eligible for LTC benefits [3].

During the 5-year exposure period from January 1, 2016, to December 31, 2020, individual health care utilization in terms of outpatient and inpatient physician care, pharmaceutical care, physiotherapy, and medical aids was assessed. Subsequently, individuals were followed for one quarter, from January 1, 2021, to March 31, 2021. In this follow-up period, we determined whether a first-time LTC need was documented. First-time LTC need was dated to the retroactive commencement of services, which often aligns with the application date. To be considered for the study population, individuals had to be aged at least 60 years at the start of the follow-up period. Any individuals with conflicting information regarding date of birth, sex, date of first-time LTC need, or date of death were identified, or whose zip code could not be mapped to regional variables were excluded. LTC need or death before the follow-up period also led to exclusion. We further required continuous insurance coverage and residence in Germany throughout the entire observation period. All individuals meeting the inclusion criteria were included in the study.

### Exposure Variables

#### *Physician Care*

Variables relating to physician care were defined in terms of care provided by general practitioners (GPs), specialists, inpatient care, screenings and vaccinations, and disease management programs (DMPs). The extent of care provided by GPs was quantified based on the number of days on which a service provided by a GP was documented during the exposure period. Similarly, the extent of specialist care was defined by the number of days with specialist services documented. Additionally, the number of different specialist groups consulted was included as a measure of the broadness of specialist care. Except for urologists, gynecologists, and medical services related to laboratory medicine, all specialist groups were considered for both specialist care variables. Urologists and gynecologists were excluded because, as a relevant part of their



XSL•FO
RenderX

services only applies to either sex and the importance of the 2 specialties with respect to LTC need prevention could differ, including them could introduce bias. To quantify the utilization of inpatient care, we determined the frequency of hospitalizations.

The screenings and vaccinations considered were cancer screenings (colon cancer screening through colonoscopy or stool tests, skin cancer screening, and prostate cancer or cervical cancer screening), a general health check-up, and the influenza vaccination. The extent of their utilization was operationalized as the number of screenings and vaccinations that were utilized according to the SHI reimbursement interval (Multimedia Appendix 1). Breast cancer screening was not considered because, until 2024, it was only reimbursed for women aged <70 years. Neither was the pneumococcal vaccination considered, as a single vaccination is recommended for the population aged ≥60 years. If a person had already been vaccinated against pneumococcal diseases before the exposure period, we would have been unable to correctly identify the vaccination.

Regarding structured disease management, we determined whether individuals participated in the established German asthma, chronic obstructive pulmonary disease (COPD), congenital heart disease, and type I or type II diabetes DMPs at any time during the exposure period. German DMPs address all patients with the respective disease rather than focusing on high-risk individuals [14] and promote evidence-based therapy and clinical pathways [15].

### *Pharmaceutical Care*

In terms of drug treatment, both polypharmacy and prescription of potentially inappropriate medications (PIMs) that can cause adverse drug events in the older population were included. Polypharmacy was defined as the prescription of 9 or more different drug substances within the same quarter [16]. The definition of PIM was based on whether at least one of the substances from a list of potentially inadequate substances specific to the German context, including analgesics, anti-inflammatory drugs, antiarrhythmic drugs, antibiotics, anticholinergic drugs, antidepressants, antihypertensive agents, neuroleptic drugs, and sedatives, was prescribed [17]. To assess the extent of polypharmacy and PIM, the number of quarters during the exposure period in which an individual's medication met the defined criteria was computed.

### *Physiotherapy*

To assess how often physiotherapy was utilized, the number of quarters with a physiotherapy prescription during the exposure period was counted.

### *Medical Aids*

All medical aids that can be reimbursed by SHI in Germany are listed in the catalog of medical aids by the National Association of Statutory Health Insurance Funds [18]. This catalog includes 38 medical aid groups, such as insoles, walking aids, or hearing aids, that were summarized into 6 categories: disease-specific aids, aids supporting self-dependence, orthopedic aids, walking aids, hearing aids, and wheelchairs, including mobility scooters. For each category, we determined whether at least one medical aid from that category was prescribed. The assignment of medical aid groups to the 6 categories is provided in Multimedia Appendix 2.

## Covariates

Comorbidities were included based on the conditions contained in the Elixhauser comorbidity index [19], typical geriatric conditions as defined by Lübke and Meinck [20], and additional diagnosis-based definitions of arthrosis, asthma, chronic obstructive pulmonary disease, and congenital heart disease developed by the AOK Research Institute [21]. The presence of these conditions was determined based on the year before the exposure period. In this way, we ensured that diagnoses were not affected by utilization behavior during the exposure period. Both inpatient and outpatient diagnoses were considered. Outpatient diagnoses were validated by requiring them to be documented as assured ("gesichert" according to the *International Statistical Classification of Diseases and Related Health Problems, 10th revision, German Modification*) in at least two quarters. Multimedia Appendix 3 includes a complete list of the conditions included.

As sociodemographic variables, age, sex, and regional variables were included. We determined the county settlement structure type with the 4 possible outcomes: metropolitan, urban, rural with agglomeration tendency, and rural as defined by the German Federal Institute for Research on Building, Urban Affairs and Spatial Development [22]. The German Index of Socioeconomic Deprivation developed at the Robert Koch Institute measures the level of socioeconomic deprivation based on administrative data on education, employment, and income [23]. It was included at zip code level and ranges from 0 (lowest deprivation) to 1 (highest deprivation).

## Statistical Analysis

Logistic regression was used to assess the association between the variables of interest and first-time LTC need. The outcome data reflect the date when LTC insurance benefits were first claimed, rather than the point of time when a person would have been first eligible, but may not yet have applied for LTC need. Therefore, and given the 3-month follow-up, logistic regression analysis was chosen over using time-to-event methods. First, a crude single exposure model was estimated by univariate regression of first-time LTC need separately on each of the variables of interest. We then controlled for the named comorbidities and sociodemographic variables in an adjusted single exposure model. Finally, all variables of interest were integrated into a single adjusted multiple exposure model, in which covariates were also controlled for.

Owing to the considerable number of variables of interest, whose association with LTC need was tested for significance, we used *P*-values adjusted for multiple testing using a max-*t* test. This test controls the probability of falsely rejecting at least one of the hypotheses in question [24]. All metric variables of interest were categorized into a null category indicating no utilization of the respective service, and 3 categories of low, medium, and high utilization by classification according to the Fisher-Jenks algorithm [25]. Rather than assuming a linear relationship, this



XSL•FO
RenderX

flexible approach enabled picking up possible nonlinear and nonmonotonic associations. To be able to estimate the association of the number of specialist billing days and specialist groups simultaneously despite their linear dependence if categorized, we introduced a binary variable displaying whether specialist care was utilized at all and combined the null category and the low utilization category of each of the billing days and specialist groups variable into a single reference category. Data preparation and statistical analysis were conducted using Oracle 19c Enterprise Edition (Oracle Corporation) and R 4.4.2 (R Foundation for Statistical Computing). Statistical significance was evaluated at a 2-sided significance level $\alpha$ of .05.

### Sensitivity Analysis

To assess whether results were driven by effects specific to the observation period and the short duration of follow-up, we first defined 3 alternative observation periods that started on April 1, 2016, July 1, 2016, and October 1, 2016, each also consisting of a 5-year exposure period and one-quarter follow-up period, and second, in another alternative specification, we extended the follow-up period to 1 year. Third, to address the potential for reverse causality, we excluded the last year from the exposure period. For this analysis, we rescaled the utilization category thresholds of each of the exposure variables to reflect the shorter exposure period (Multimedia Appendix 4).

### Ethical Considerations

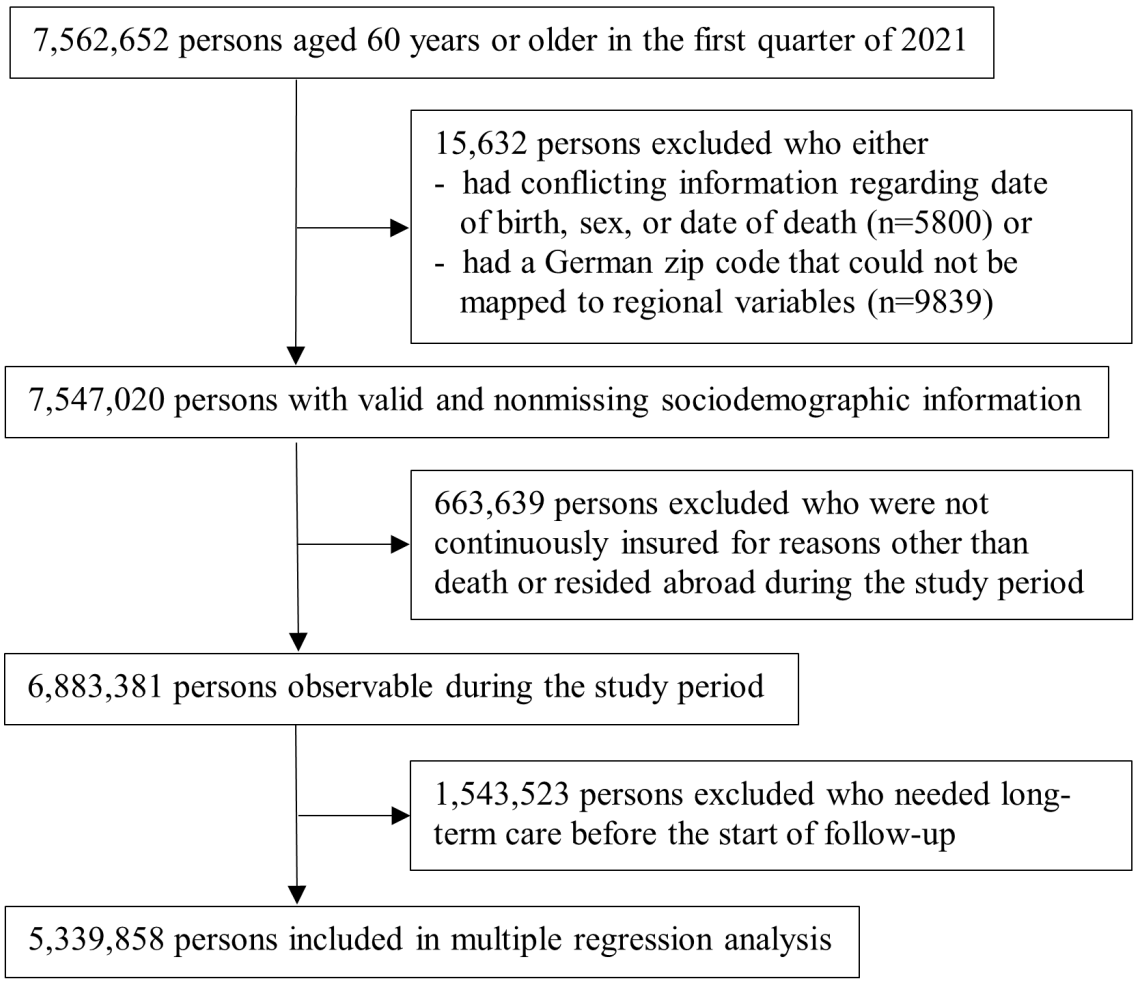
All analyses described were conducted with the approval of the responsible ethics committee of Charité–Universitätsmedizin Berlin (EA4/147/23), in accordance with national law and the current version of the Declaration of Helsinki. We report our findings in accordance with the STROBE (Strengthening of the Reporting of Observational Studies in Epidemiology) guidelines.

## Results

### Description of the Study Population

Figure 1 shows the flowchart of study population selection. At the start of the follow-up, 7,562,652 persons aged ≥60 years were insured with AOK. After application of the selection criteria, the final study population comprised 5,339,858 insured persons. Descriptive statistics for the study population are presented in Table 1. The average age was 71.3 (SD 8.16) years, and 54.3% (n=2,899,228) were women. There were no missing values. During the one-quarter follow-up, 91,160 (1.7%) individuals became in need of LTC. See Multimedia Appendices 3-7 for flowcharts and full descriptive statistics, including the sensitivity analysis study populations.

**Figure 1.** Flowchart of study population selection.






XSL•FO
RenderX

**Table .** Descriptive statistics for the study population (n=5,339,858).

| Characteristics | Value |
|---|---|
| Age (y), mean (SD) | 71.3 (8.2) |
| Sex, female, n (%) | 2,899,228 (54.3) |
| German Index of Social Deprivation, mean (SD) | 0.5 (0.2) |
| County settlement structure type, n (%) | |
| Metropolitan | 1,136,388 (21.3) |
| Urban | 1,976,750 (37.0) |
| Rural with agglomeration tendency | 1,186,650 (22.2) |
| Rural | 1,040,070 (19.4) |
| Utilization of general practitioner (number of days), n (%) | |
| None (0) | 118,747 (2.2) |
| Low (1-47) | 2,568,788 (48.1) |
| Medium (48-98) | 2,159,401 (40.4) |
| High (>98) | 492,922 (9.2) |
| Utilization of specialist, n (%) | 5,053,869 (94.6) |
| Utilization of specialist (number of groups), n (%) | |
| Low (0 - 4) | 2,415,753 (45.2) |
| Medium (5-7) | 1,846,289 (34.6) |
| High (>7) | 1,077,816 (20.2) |
| Utilization of specialist (number of days), n (%) | |
| Low (0 - 47) | 4,259,918 (79.8) |
| Medium (48-396) | 1,075,012 (20.1) |
| High (>396) | 4928 (0.1) |
| Hospitalizations, n (%) | |
| None (0) | 2,594,401 (48.6) |
| Low (1-2) | 1,889,121 (35.4) |
| Medium (3-6) | 734,973 (13.8) |
| High (>6) | 121,363 (2.3) |
| Screenings and vaccinations (number of services), n (%) | |
| None (0) | 852,863 (16.0) |
| Low (1) | 1,704,527 (31.9) |
| Medium (2) | 1,513,223 (28.3) |
| High (>2) | 1,269,245 (23.8) |
| DMP[a] congenital heart disease enrollment, n (%) | 531,093 (9.9) |
| DMP chronic obstructive pulmonary disease enrollment, n (%) | 218,486 (4.1) |
| DMP asthma enrollment, n (%) | 146,291 (2.7) |
| DMP diabetes enrollment, n (%) | 1,113,656 (20.9) |
| Polypharmacy (number of quarters), n (%) | |
| None (0) | 4,425,424 (82.9) |
| Low (1-4) | 617,878 (11.6) |
| Medium (5-11) | 193,477 (3.6) |
| High (>11) | 103,079 (1.9) |
| Prescription of potentially inadequate medications (number of quarters), n (%) | |



XSL•FO
RenderX

| Characteristics | Value |
|---|---|
| None (0) | 3,669,247 (68.7) |
| Low (1-4) | 1,136,815 (21.3) |
| Medium (5-12) | 282,097 (5.3) |
| High (>12) | 251,699 (4.7) |
| Physiotherapy (number of quarters), n (%) | |
| None (0) | 2,559,641 (47.9) |
| Low (1-4) | 2,005,639 (37.6) |
| Medium (5-10) | 578,229 (10.8) |
| High (>10) | 196,349 (3.7) |
| Orthopedic aids prescription, n (%) | 2,259,257 (42.3) |
| Hearing aids prescription, n (%) | 567,833 (10.6) |
| Walking aids prescription, n (%) | 593,506 (11.1) |
| Wheelchairs including mobility scooters prescription, n (%) | 57,821 (1.1) |
| Aids supporting self-dependence prescription, n (%) | 853,723 (16.0) |
| Disease-specific aids prescription, n (%) | 1,622,125 (30.4) |

[a]Disease management program.

## Regression Analysis

The odds ratios (ORs) resulting from estimating association of health care utilization with first-time need for LTC are presented in Table 2. The results reported in the following sections refer to the adjusted multiple exposure model.



XSL•FO
RenderX

**Table .** Results of logistic regression of first-time long-term care need on health care utilization variables.

| Variables | Crude single exposure model | | Adjusted single exposure model[a] | | Adjusted multiple exposure model[a] | |
|---|---|---|---|---|---|---|
| | Odds ratio (95% CI) | *P* value | Odds ratio (95% CI) | *P* value | Odds ratio (95% CI) | *P* value |
| Utilization of general practitioner (number of billing days) | | | | | | |
| None (0) | Reference | —[b] | Reference | — | Reference | — |
| Low (1-47) | 1.28 (1.20 - 1.37) | <.001 | 1.26 (1.18 - 1.34) | <.001 | 1.38 (1.29 - 1.48) | <.001 |
| Medium (48-98) | 2.65 (2.49 - 2.83) | <.001 | 1.65 (1.55 - 1.76) | <.001 | 1.59 (1.48 - 1.71) | <.001 |
| High (>98) | 4.35 (4.08 - 4.65) | <.001 | 2.16 (2.02 - 2.31) | <.001 | 1.65 (1.53 - 1.77) | <.001 |
| Utilization of specialist | | | | | | |
| None (0 groups and billing days) | Reference | — | Reference | — | Reference | — |
| Positive number of groups (>0) and billing days (>0) | 1.12 (1.08 - 1.16) | <.001 | 0.95 (0.92 - 0.98) | <.001 | 0.82 (0.79 - 0.85) | <.001 |
| Utilization of specialist (number of specialist groups) | | | | | | |
| Low (0 - 4) | Reference | — | Reference | — | Reference | — |
| Medium (5-7) | 1.03 (1.02 - 1.05) | <.001 | 1.03 (1.02 - 1.05) | <.001 | 0.95 (0.93 - 0.96) | <.001 |
| High (>7) | 1.02 (1.00 - 1.04) | .07 | 1.12 (1.09 - 1.14) | <.001 | 0.94 (0.92 - 0.96) | <.001 |
| Utilization of specialist (number of billing days) | | | | | | |
| Low (0 - 47) | Reference | — | Reference | — | Reference | — |
| Medium (48-396) | 1.52 (1.50 - 1.55) | <.001 | 1.20 (1.18 - 1.22) | <.001 | 1.06 (1.04 - 1.08) | <.001 |
| High (>396) | 3.37 (2.96 - 3.82) | <.001 | 3.09 (2.70 - 3.53) | <.001 | 1.70 (1.48 - 1.95) | <.001 |
| Hospitalizations | | | | | | |
| None (0) | Reference | — | Reference | — | Reference | — |
| Low (1-2) | 1.96 (1.93 - 1.99) | <.001 | 1.62 (1.59 - 1.65) | <.001 | 1.46 (1.44 - 1.49) | <.001 |
| Medium (3-6) | 3.47 (3.41 - 3.53) | <.001 | 2.55 (2.50 - 2.60) | <.001 | 2.00 (1.96 - 2.05) | <.001 |
| High (>6) | 5.98 (5.82 - 6.15) | <.001 | 4.53 (4.40 - 4.67) | <.001 | 3.07 (2.97 - 3.17) | <.001 |
| Screenings and vaccinations (number of services) | | | | | | |
| None (0) | Reference | — | Reference | — | Reference | — |
| Low (1) | 1.11 (1.09 - 1.13) | <.001 | 0.94 (0.93 - 0.96) | <.001 | 0.90 (0.88 - 0.92) | <.001 |
| Medium (2) | 1.08 (1.06 - 1.10) | <.001 | 0.84 (0.82 - 0.86) | <.001 | 0.77 (0.76 - 0.79) | <.001 |
| High (>2) | 0.91 (0.89 - 0.93) | <.001 | 0.68 (0.66 - 0.69) | <.001 | 0.62 (0.60 - 0.64) | <.001 |
| DMP[c] congenital heart disease enrollment | 1.74 (1.70 - 1.77) | <.001 | 0.98 (0.95 - 1.00) | .09 | 0.93 (0.91 - 0.96) | <.001 |
| DMP chronic obstructive pulmonary disease enrollment | 1.66 (1.62 - 1.70) | <.001 | 1.24 (1.19 - 1.28) | <.001 | 1.19 (1.14 - 1.23) | <.001 |



XSL•FO
RenderX

| Variables | Crude single exposure model | | Adjusted single exposure model[a] | | Adjusted multiple exposure model[a] | |
|---|---|---|---|---|---|---|
| | Odds ratio (95% CI) | *P* value | Odds ratio (95% CI) | *P* value | Odds ratio (95% CI) | *P* value |
| DMP asthma enrollment | 1.05 (1.01 - 1.09) | .03 | 0.96 (0.91 - 1.01) | .11 | 0.98 (0.93 - 1.03) | >.99 |
| DMP diabetes enrollment | 1.68 (1.65 - 1.70) | <.001 | 1.06 (1.03 - 1.09) | <.001 | 0.99 (0.96 - 1.02) | >.99 |
| Polypharmacy (number of quarters) | | | | | | |
| None (0) | Reference | — | Reference | — | Reference | — |
| Low (1-4) | 2.58 (2.54 - 2.62) | <.001 | 1.88 (1.85 - 1.91) | <.001 | 1.38 (1.35 - 1.40) | <.001 |
| Medium (5-11) | 3.37 (3.29 - 3.45) | <.001 | 2.30 (2.24 - 2.36) | <.001 | 1.49 (1.45 - 1.53) | <.001 |
| High (>11) | 4.09 (3.97 - 4.21) | <.001 | 2.84 (2.75 - 2.94) | <.001 | 1.68 (1.62 - 1.74) | <.001 |
| Prescription of potentially inadequate medications (number of quarters) | | | | | | |
| None (0) | Reference | — | Reference | — | Reference | — |
| Low (1-4) | 1.38 (1.36 - 1.40) | <.001 | 1.27 (1.25 - 1.29) | <.001 | 1.08 (1.07 - 1.10) | <.001 |
| Medium (5-12) | 1.86 (1.81 - 1.91) | <.001 | 1.44 (1.40 - 1.47) | <.001 | 1.11 (1.08 - 1.14) | <.001 |
| High (>12) | 2.18 (2.13 - 2.24) | <.001 | 1.42 (1.39 - 1.46) | <.001 | 1.14 (1.11 - 1.17) | <.001 |
| Physiotherapy (number of quarters) | | | | | | |
| None (0) | Reference | — | Reference | — | Reference | — |
| Low (1-4) | 1.00 (0.99 - 1.01) | >.99 | 1.01 (1.00 - 1.03) | 0.11 | 0.85 (0.84 - 0.86) | <.001 |
| Medium (5-10) | 1.16 (1.14 - 1.19) | <.001 | 1.13 (1.10 - 1.15) | <.001 | 0.81 (0.79 - 0.83) | <.001 |
| High (>10) | 1.61 (1.57 - 1.66) | <.001 | 1.42 (1.37 - 1.46) | <.001 | 0.98 (0.95 - 1.01) | >.99 |
| Orthopedic aids prescription | 1.09 (1.07 - 1.10) | <.001 | 1.11 (1.09 - 1.12) | <.001 | 0.94 (0.93 - 0.96) | <.001 |
| Hearing aids prescription | 2.01 (1.97 - 2.04) | <.001 | 1.01 (0.99 - 1.02) | 0.44 | 0.98 (0.96 - 0.99) | 0.22 |
| Walking aids prescription | 3.66 (3.61 - 3.71) | <.001 | 2.13 (2.09 - 2.16) | <.001 | 1.62 (1.59 - 1.64) | <.001 |
| Wheelchairs including mobility scooters prescription | 3.84 (3.71 - 3.98) | <.001 | 2.38 (2.30 - 2.47) | <.001 | 1.43 (1.38 - 1.49) | <.001 |
| Aids supporting self-dependence prescription | 2.84 (2.80 - 2.88) | <.001 | 1.80 (1.77 - 1.83) | <.001 | 1.31 (1.29 - 1.33) | <.001 |
| Disease-specific aids prescription | 1.67 (1.65 - 1.70) | <.001 | 1.43 (1.41 - 1.45) | <.001 | 1.14 (1.12 - 1.16) | <.001 |

[a]Adjusted for comorbidities, age, sex, county settlement structure type, and German Index of Social Deprivation.

[b]Not applicable.

[c]DMP: disease management program.

### *Physician Care*

Regarding outpatient physician care, a low number of GP visits was related to a significant increase in the odds of first-time LTC need (OR 1.38, 95% CI 1.29 - 1.48; *P*<.001) relative to none. Odds for first-time LTC need rose further in medium (OR 1.59, 95% CI 1.48 - 1.71; *P*<.001) and high levels of utilization (OR 1.65, 95% CI 1.53 - 1.77; *P*<.001). With a low number of specialist groups and billing days, utilization of specialist care entailed a reduction of odds versus no utilization (OR 0.82, 95% CI 0.79 - 0.85; *P*<.001). Compared to this low utilization level,



XSL•FO
RenderX

odds reduced further in medium (OR 0.95, 95% CI 0.93 - 0.96; *P*<.001) and high (OR 0.94, 95% CI 0.92 - 0.96; *P*<.001) numbers of specialist groups but increased in medium (OR 1.06, 95% CI 1.04 - 1.08; *P*<.001) and high (OR 1.70, 95% CI 1.48 - 1.95; *P*<.001) numbers of specialist days. Hospitalizations were strongly and significantly associated with a higher risk of first-time LTC need. More than 6 hospitalizations within the 5-year exposure period were related to a more than 3-fold increase in the odds of first-time LTC need (OR 3.07, 95% CI 2.97 - 3.17; *P*<.001) relative to none.

A high utilization of screenings and vaccinations was related to a reduction in the odds of first-time LTC need (OR 0.62, 95% CI 0.60 - 0.64; *P*<.001). Although less strong in magnitude, low (OR 0.90, 95% CI 0.88 - 0.92; *P*<.001) and medium (OR 0.62, 95% CI 0.60 - 0.64; *P*<.001) utilization of screenings and vaccinations were also significantly negatively associated with first-time LTC need. Of the 4 DMPs, only the congenital heart disease DMP exhibited a significant protective relationship with first-time LTC need (OR 0.93, 95% CI 0.91 - 0.96; *P*<.001). COPD DMP participation predicted an increased risk of first-time LTC need (OR 1.19, 95% CI 1.14 - 1.23; *P*<.001).

### *Pharmaceutical Care*

Both prescription of PIM and polypharmacy were accompanied by higher odds of first-time LTC need. The relationship is stronger in magnitude for high levels of polypharmacy (OR 1.68, 95% CI 1.62 - 1.74; *P*<.001) than for high levels of PIM (OR 1.14, 95% CI 1.11 - 1.17; *P*<.001).

### *Physiotherapy*

Relative to nonuse, a low (OR 0.85, 95% CI 0.84 - 0.86; *P*<.001) or medium (OR 0.81, 95% CI 0.79 - 0.83; *P*<.001) utilization of physiotherapy correlated negatively with first-time LTC need. With a high utilization level, the association did not remain statistically significant (*P*>.99).

### *Medical Aids*

Orthopedic aids had a significant negative association with first-time LTC need (OR 0.94, 95% CI 0.93 - 0.96; *P*<.001). Walking aids (OR 1.62, 95% CI 1.59 - 1.64; *P*<.001), wheelchairs including mobility scooters (OR 1.43, 95% CI 1.38 - 1.49; *P*<.001), aids supporting self-dependence (OR 1.31, 95% CI 1.29 - 1.33; *P*<.001), and disease-specific aids (OR 1.14, 95% CI 1.12 - 1.16; *P*<.001) were related to an increased risk of first-time LTC need.

## Sensitivity Analysis

As shown in Multimedia Appendix 8, the 3 alternative observation periods yielded similar results compared to the main analysis. Multimedia Appendix 9 presents the results obtained when extending the follow-up period to 1 year. While most associations exhibited no notable difference to the main analysis, the relationship between LTC need and the number of hospitalizations became weaker, and the negative associations of the DMPs for asthma (OR 0.95, 95% CI 0.92 - 0.98; *P*=.008) and diabetes (OR 0.95, 95% CI 0.93 - 0.96; *P*<.001) with LTC need became significant. Omitting the final year of the exposure period, as displayed in Multimedia Appendix 10, produced analogous estimates but for the number of hospitalizations. As when extending the follow-up duration to 1 year, hospitalizations displayed lower estimates for all utilization categories. Relative to the main analysis, the OR of more than 6 hospitalizations (OR 1.80, 95% CI 1.74 - 1.86; *P*<.001) was almost halved.

# Discussion

## Principal Findings

We have shown that a wide range of measures of health care utilization display significant and strong associations with the risk of first-time LTC need. To the best of our knowledge, our study is the first to analyze the interplay of the frequency of physician visits, DMP enrollment, physiotherapy utilization, prescription of medical aids, screenings and vaccinations, and inadequate medication supply with first-time LTC need simultaneously. In the population aged ≥60 years, utilization of specialist care, physiotherapy, and screenings and vaccinations display the strongest relationship with nonoccurrence of first-time LTC need. At the same time, frequent utilization of GP care, hospitalizations, polypharmacy, and prescription of walking aids showed a significant association with first-time LTC need. As this study relied on a 5-year exposure period to evaluate health care utilization, the relationship between health care use in a 5-year period and LTC need can only be assessed in and does only apply to individuals who neither need LTC nor die within 5 years. This approach provides a longer-term perspective on the health care trajectory of a person, which may be more meaningful for prevention than a single event.

## Comparison With Prior Work

Apart from the 2 studies, which corroborated that inadequate medication supply is related to an increase in future LTC risk [10,11], further evidence on the relationship of health care utilization with first-time LTC need is missing. LTC need being a sociolegal concept in Germany that differs substantially from that in other countries [3,26], prior research has focused on related outcomes, such as frailty, functional measures, morbidity, or mortality [8,27,28]. Although these outcomes must be distinguished both from each other and first-time LTC need, they are interrelated [29-32] and evidence on how they correlate with health care utilization can promote our understanding of the link between health care utilization and first-time LTC need.

### *Frequency of Physician Contact*

We evaluated 3 variables that are closely related to the frequency of physician contact: GP billing days, specialist physician billing days, and hospital admissions. Compared to no contact, both GP billing days and hospital admissions were related to an increased risk of first-time LTC need. While utilization of specialist care with a low number of billing days and specialist groups showed a preventive association, all 3 variables have in common that compared to a low level of utilization, a higher level of physician contact was related to an increase in the risk of first-time LTC need. On the one hand, frequent physician contact may entail constant assessment of health risks and subsequent counseling and thereby exerting a preventive effect on LTC need risk [33,34]. On the other hand, there is a wealth of literature on how physical frailty is associated with a higher utilization of health care in terms of outpatient care, inpatient



XSL•FO
RenderX

care, home care, and nursing home admission [35-40]. Moreover, individuals with an adverse status in terms of living alone, social relations, and social support, referred to as socially frail, are reported to have a lower number of physician visits and lower likelihood of hospitalization, hinting at lower ability to access health care services including LTC [41]. Despite our efforts to control for comorbidities, both physical and social frailty likely act as confounders, correlating both with first-time LTC need risk and health care utilization.

In sensitivity analyses that excluded the last year from the exposure period or extended follow-up to 1 year, we also observed a markedly weaker association between hospitalizations and LTC need. This is consistent with the fact that hospitalizations frequently occur shortly before first-time LTC need [42]. If there is a time lag between the end of the exposure period and the date of first-time LTC need, hospitalizations shortly before first-time LTC need will not be reflected in the exposure variable. This underscores how the strong relationship between hospitalizations and LTC need in the main analysis reflects morbidity levels. In summary, while the frequency of physician contact is a valuable identifier of persons at risk of first-time LTC need, it cannot be assumed to be detrimental to independent living based on our analysis.

### *Condition-Specific Health Care Utilization*

Apart from the frequency of physician contact, we also evaluated how extensive specialist outpatient care utilization was by including the number of specialist groups consulted. At a constant number of specialist billing days, an increase in specialist groups consulted was connected to a decreased risk of first-time LTC need. Similarly, another claims data–based study from Germany reported that, adjusted for GP care utilization, LTC beneficiaries are less likely to visit physicians from different specialist groups than older people without LTC need [43]. We cannot rule out that part of the promising association we estimated pertains to barriers to accessibility of specialist care. These might be more pronounced for individuals at higher risk of first-time LTC need. Still, hypothesizing that even with few patient-physician contacts, each specialist group can give specific preventive recommendations and guidance, it is plausible that a wider variety of specialist groups consulted, holding the number of billing days fixed, exerts a preventive effect. On the other hand, having visited a given specialist more frequently may reflect higher morbidity levels more so than an improved capacity to provide preventive guidance.

Of the 4 DMPs evaluated, we found a preventive relationship consistent across main and sensitivity analyses regarding only the congenital heart disease DMP, while the COPD DMP even showed a detrimental association. As we can only account for disease presence and not severity, selection effects might bias our estimates if DMP participants differ from nonparticipants in this regard. DMPs have not yet been evaluated in terms of their association with first-time LTC need, nor have there been DMP evaluations using randomized control trials in Germany [44].

While we found physiotherapy utilization to be predictive of nonoccurrence of first-time LTC need, when receiving prescriptions in up to 10 quarters within 5 years, we saw no such association for higher levels of utilization. There is prior evidence that physiotherapy can support the use of medical aids [45], prevent institutionalization after hospitalization [46], and, in the case of a Spanish study, reduce mortality through improvement in quality of life, alleviation of musculoskeletal disorders, a referral function in the health care system, and fall prevention [47]. While there is currently no evidence on the relationship between physiotherapy utilization and first-time LTC need, 2 articles report that frail persons receive physiotherapy more often than nonfrail ones [36,38].

Furthermore, our medical aids analysis showed that only orthopedic devices had a negative relationship with first-time LTC need. All other categories except for hearing aids had a positive association that was most pronounced in walking aids. A meta-analysis suggested that insoles, which make up for the largest share of prescriptions within the orthopedic devices group, were beneficial for balance improvement and fall prevention [48], supporting our finding of a preventive potential. Although walking aids are intended to prevent falls, walking aid users face a high risk of falls and injuries that might be reduced by improvements in devices and better instructions on device usage [45,49]. Likely, both difficulties in accounting for confounders such as gait quality and potential for improvement in walking aid use contribute to the strong positive association with first-time LTC need we observe.

### *Screenings and Vaccinations*

The cancer screenings, general health check, and influenza vaccination summarized in a single variable displayed a strong relationship with nonoccurrence of first-time LTC need. There is tentative evidence supporting the health benefits of these services in the population aged ≥60 years, yet the association we find remains high against this backdrop.

Despite the potential of cancer screening to reduce all-cause mortality by 1% to 3% in Germany [50], screening cessation may be recommended for prostate cancer, cervical cancer, and colorectal cancer. With increasing age and decreasing life expectancy, potential harms in screening may outweigh potential benefits, necessitating an individualized screening decision [51-53]. On the one hand, this calls into question whether cancer screenings can have a substantial preventive effect in terms of LTC need. On the other hand, it raises the question of whether part of the observed effect reflects screening cessation recommendations being put into practice. General health checks and health risk assessments have been shown to increase uptake of other preventive services, but an effect on mortality cannot be substantiated despite positive impact on chronic diseases treatment and patient-reported outcomes [34,54-56]. The influenza vaccination provides the most concrete reference for a potential delaying effect on first-time LTC need. Not only is it crucial to influenza prevention, especially so in older adults [57], but it is also related to a decreased risk of dementia [58,59], the third most-named diagnosis for first-time LTC need in Germany in 2021 [60].

Other than effects of screenings and vaccinations themselves, preventive health behavior could contribute to the strength of the observed association. It has been shown that the uptake of preventive services is higher in individuals with



XSL•FO
RenderX

health-promoting behavior [61,62]. As we have no information on crucial behavioral aspects such as physical activity, nutrition, and alcohol and tobacco use, they likely bias the observed relationship toward prevention. Bearing that in mind, our study provides first hints for a possible delaying effect of screenings and vaccinations on first-time LTC need that further studies should investigate in more detail on a broader base of data.

### *Inadequate Medication Supply*

Our finding that inadequate medication supply is positively related to first-time LTC need and polypharmacy, more so than PIM, is in line with prior cross-sectional [63] and longitudinal studies [10,11] from Japan. Kuroda et al [10] evaluated polypharmacy and PIM using drug prescriptions from a 2-year exposure period and reported associations of similar magnitude. Inadequate medication supply has been shown to be associated with falls, dementia, functional and cognitive decline, and sarcopenia [28,64,65], all of which are plausible causes of first-time LTC need. This suggests that correction of inadequate medication supply can be a leverage point to LTC need prevention.

## Strengths and Limitations

This retrospective cohort study adheres to certain conditions necessary to justify causal statements about the observed associations. Outcome was evaluated after exposure, and thus, reverse causality through effects of utilization of LTC insurance services can be ruled out. Still, utilization might be driven by morbidity trends on a trajectory approaching first-time LTC need [42]. Excluding the final year of the exposure period from the analysis showed that only estimates related to hospitalization are affected strongly by such morbidity trends. The robustness of the remaining estimates indicates that reverse causality through short-term morbidity trends is no major concern. The large nationwide sample enabled us to assess utilization of multiple health care services simultaneously despite their interrelation. In this way, we were able to measure the relationship of each utilization variable with first-time LTC need given the utilization pattern in terms of the other variables.

Nonetheless, this study faces some substantial limitations that confine the results to the domain of correlation. It is the nature of claims data that they lack information on disease severity, health-promoting behavior, social networks, and individual-level socioeconomic variables. These parameters need to be integrated to enable attainment of unbiased estimates, given that they are related to physical and social frailty, as well as the accessibility of health care [66,67]. Furthermore, despite the large sample size, results attained based on AOK-insured individuals cannot be generalized to the German population, as they tend to have lower socioeconomic status and higher comorbidity burden than people insured privately or with other SHI funds [68,69]. On top of that, our observation period coincides with the COVID-19 pandemic, whose effect on health care utilization limits the external temporal validity of our study.

Moreover, a range of health care services that may contribute to LTC need prevention, such as breast cancer screening, geriatric assessment, and rehabilitation, were not evaluated as part of this study. This study was designed to estimate associations of health care utilization with first-time LTC need without placing further restrictions on the study population than the criterion of being aged ≥60 years. Therefore, services such as breast cancer screening and geriatric assessment, access to which is specific to sex and age, could not be included. In Germany, rehabilitation services are covered by the statutory pension insurance scheme before retirement. Utilization of rehabilitation services can thus not be assessed on SHI claims data alone for a population partially still in employment. Furthermore, we primarily aimed to assess the association of health care utilization with first-time LTC need as a proxy for functional health. We must concede that claiming LTC benefits is itself an act of health care utilization that, besides functional health, is determined by access to health care and utilization behavior. As a determinant of health care accessibility and utilization behavior, individual-level socioeconomic variables could be a valuable addition in factoring out its confounding impact.

## Conclusions

In conclusion, we have shown that the utilization of recommended screenings, vaccinations, specialist physician care, and physiotherapy displays substantial and significant associations with nonoccurrence of first-time LTC need in the German population aged ≥60 years. Our study provides an important foundation for future research. It is essential to address the limitations posed by unobserved confounding variables and the nonexperimental study design to generate evidence on how LTC need can be delayed or avoided. Such evidence could provide guidance on orienting health care toward preventing functional decline.

## Acknowledgments

We thank Dr Antje Schwinger for her contributions during the initial phase of the study and Ms Hanna Müller for her assistance with literature review.

## Funding

This study has been funded by the Federal Joint Committee (G-BA) as part of the WEGE project (grant ID 01VSF22047). The funder had no involvement in the study design, data collection, analysis, interpretation, or the writing of the manuscript.



XSL•FO
RenderX

## Data Availability

The authors confirm that the statutory health insurance claims data used in this study cannot be made available in the manuscript, the supplemental files, or a public repository due to German data protection laws ("Bundesdatenschutzgesetz"). Therefore, AOK claims data are stored on a secure drive in the AOK Research Institute to facilitate replication of the results. Generally, access to data of statutory health insurance funds for research purposes is possible only under the conditions defined in German Social Law (SGB V § 287). Requests for data access can be sent as a formal proposal specifying the recipient and purpose of the data transfer to the appropriate data protection agency. Access to the data used in this study can only be provided to external parties under the conditions of the cooperation contract of this research project and after written approval by the SHI fund. For assistance in obtaining access to AOK claims data, please contact wido@wido.bv.aok.de.

## Authors' Contributions

Conceptualization: TB (equal), KJ (equal), CH (equal), S Behrendt (equal), S Blüher (equal), SS (equal), AS (equal)
Data curation: TB (equal), KJ (equal)
Formal analysis: TB
Funding acquisition: S Blüher (lead), AS (equal)
Methodology: TB (lead), KJ (supporting), SS (supporting)
Project administration: CH (equal), S Blüher (equal)
Software: TB
Supervision: SS (lead), S Blüher (equal)
Visualization: TB
Writing – original draft preparation: TB
Writing – review and editing: TB (lead), KJ (equal), CH (equal), S Behrendt (equal), S Blüher (equal), SS (equal).

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Definition of the screenings and vaccinations utilization variable.
[DOCX File, 19 KB - i-jmr_v15i1e86572_app1.docx ]

Multimedia Appendix 2
Assignment of medical aid groups to medical aid categories.
[DOCX File, 23 KB - i-jmr_v15i1e86572_app2.docx ]

Multimedia Appendix 3
Descriptive statistics of study population in main analysis and sensitivity analyses 1 to 3.
[DOCX File, 65 KB - i-jmr_v15i1e86572_app3.docx ]

Multimedia Appendix 4
Descriptive statistics for study population in sensitivity analysis 5.
[DOCX File, 32 KB - i-jmr_v15i1e86572_app4.docx ]

Multimedia Appendix 5
Flowchart of study population selection for sensitivity analysis 1.
[DOCX File, 24 KB - i-jmr_v15i1e86572_app5.docx ]

Multimedia Appendix 6
Flowchart of study population selection for sensitivity analysis 2.
[DOCX File, 24 KB - i-jmr_v15i1e86572_app6.docx ]

Multimedia Appendix 7
Flowchart of study population selection for sensitivity analysis 3.
[DOCX File, 24 KB - i-jmr_v15i1e86572_app7.docx ]

Multimedia Appendix 8



XSL•FO
RenderX

Results of logistic regression of first-time long-term care need on health care utilization variables in main analysis and sensitivity analyses 1 to 3.
[DOCX File, 87 KB - i-jmr_v15i1e86572_app8.docx ]

Multimedia Appendix 9
Results of logistic regression of first-time long-term care need on health care utilization variables in main analysis and sensitivity analysis 4.
[DOCX File, 70 KB - i-jmr_v15i1e86572_app9.docx ]

Multimedia Appendix 10
Results of logistic regression of first-time long-term care need on health care utilization variables in main analysis and sensitivity analysis 5.
[DOCX File, 68 KB - i-jmr_v15i1e86572_app10.docx ]

## References


1. Park S, Stimpson JP, Fendrick AM. Health care utilization patterns among adults with or without functional disabilities. JAMA Netw Open 2025 Apr 1;8(4):e254729. [doi: 10.1001/jamanetworkopen.2025.4729] [Medline: 40214987]
2. Press release no.478. Federal Statistical Office of Germany. 2024. URL: https://www.destatis.de/DE/Presse/Pressemitteilungen/2024/12/PD24_478_224.html [accessed 2025-08-11]
3. Geyer J, Börsch-Supan AH, Haan P, Perdrix E. Long-term care in Germany. : National Bureau of Economic Research Working Paper Series; 2023. [doi: 10.3386/w31870]
4. Zahl der pflegebedürftigen in deutschland steigt überraschend massiv. Deutsches Ärzteblatt. 2024. URL: https://www.aerzteblatt.de/news/zahl-der-pflegebeduerftigen-in-deutschland-steigt-ueberraschend-massiv-26638455-f8c6-4239-a35c-c42e2c0912c4 [accessed 2025-08-11]
5. United Nations Department of Economic and Social Affairs. World Social Report 2023. : United Nations; 2023 Jan 12. [doi: 10.18356/9789210019682]
6. Pflegevorausberechnung - deutschland und bundesländer, berichtszeitraum 2022-2070. Federal Statistical Office of Germany. 2023. URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsvorausberechnung/Publikationen/Downloads-Vorausberechnung/statistischer-bericht-pflegevorausberechnung-5124209229005.html [accessed 2025-10-13]
7. 15 koordinierte bevölkerungsvorausberechnung - deutschland. Federal Statistical Office of Germany. 2022. URL: https://www-genesis.destatis.de/datenbank/online/statistic/12421/details [accessed 2025-10-13]
8. Apóstolo J, Cooke R, Bobrowicz-Campos E, et al. Effectiveness of interventions to prevent pre-frailty and frailty progression in older adults: a systematic review. JBI Database System Rev Implement Rep 2018;16(1):140-232. [doi: 10.11124/JBISRIR-2017-003382]
9. Schneider A, Blüher S, Grittner U, et al. Is there an association between social determinants and care dependency risk? A multi‐state model analysis of a longitudinal study. Res Nurs Health 2020 Jun;43(3):230-240. [doi: 10.1002/nur.22022]
10. Kuroda N, Iwagami M, Hamada S, Komiyama J, Mori T, Tamiya N. Associations of polypharmacy and drugs with sedative or anticholinergic properties with the risk of long‐term care needs certification among older adults in Japan: A population‐based, nested case–control study. Geriatrics Gerontology Int 2022 Jul;22(7):497-504. [doi: 10.1111/ggi.14393]
11. Maeda K, Kawashiri SY, Arima K, et al. Association between polypharmacy and the risk of requiring long-term care among community-dwelling older residents: a retrospective cohort study. J Gen Fam Med 2025 Sep;26(5):402-407. [doi: 10.1002/jgf2.70041] [Medline: 40904436]
12. Hashimoto A, Hashimoto H, Kawaguchi H. Technical efficiency of prevention services for functional dependency in Japan's public long-term care insurance system: an ecological study. Int J Health Policy Manag 2024;13(1):8226. [doi: 10.34172/ijhpm.8226] [Medline: 39099485]
13. Zahlen und fakten kompakt 2025. AOK. 2025. URL: https://www.aok.de/pp/zahlen-und-fakten/das-wichtigste-in-kuerze/ [accessed 2026-04-15]
14. Burns J, Kurz C, Laxy M. Effectiveness of the German disease management programs: quasi-experimental analyses assessing the population-level health impact. BMC Public Health 2021 Nov 15;21(1):2092. [doi: 10.1186/s12889-021-12050-7] [Medline: 34781907]
15. Busse R. Disease management programs In Germany's statutory health insurance system. Health Aff 2004 May;23(3):56-67. [doi: 10.1377/hlthaff.23.3.56]
16. Masnoon N, Shakib S, Kalisch-Ellett L, Caughey GE. What is polypharmacy? A systematic review of definitions. BMC Geriatr 2017 Oct 10;17(1):230. [doi: 10.1186/s12877-017-0621-2] [Medline: 29017448]
17. Holt S, Schmiedl S, Thürmann PA. Potentially inappropriate medications in the elderly: the PRISCUS list. Dtsch Arztebl Int 2010 Aug;107(31-32):543-551. [doi: 10.3238/arztebl.2010.0543] [Medline: 20827352]




XSL•FO
RenderX

18. Medical aids list. National Association of Statutory Health Insurance Funds. 2025. URL: https://hilfsmittel.gkv-spitzenverband.de/home [accessed 2025-08-11]
19. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005 Nov;43(11):1130-1139. [doi: 10.1097/01.mlr.0000182534.19832.83] [Medline: 16224307]
20. Lübke N, Meinck M. Geriatric multimorbidity in claims data - part 1. Analysis of hospital data and long-term care insurance data. Z Gerontol Geriatr 2012 Aug;45(6):485-497. [doi: 10.1007/s00391-012-0301-y] [Medline: 22538783]
21. Schlotmann A, Weirauch H, Brückner G. Gesundheitsatlas-deutschland.de: Methodik. Gesundheitsatlas Deutschland. 2025. URL: https://www.gesundheitsatlas-deutschland.de/data/Downloads/gesundheitsatlas_deutschland_methodik.pdf [accessed 2025-09-04]
22. Markus B, Markus E, Petra K. Raumabgrenzungen Und Raumtypen Des BBSR: Federal Institute for Research on Building, Urban Affairs, and Spatial Development; 2012.
23. Michalski N, Reis M, Tetzlaff F, et al. German Index of Socioeconomic Deprivation (GISD): revision, update and applications. J Health Monit 2022 Dec;7(Suppl 5):2-23. [doi: 10.25646/10641] [Medline: 36628258]
24. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. Ludwig-Maximilians-Universität München. 2008. URL: https://epub.ub.uni-muenchen.de/2120/ [accessed 2025-08-11]
25. Slocum TA, McMaster RB, Kessler FC, Howard HH. Thematic Cartography and Geographic Visualization: Prentice Hall; 2005.
26. Llena-Nozal A, Barszczewski J, Rauet-Tejeda J. How do countries compare in their design of long-term care provision?: a typology of long-term care systems. : OECD Health Working Papers; 2025. [doi: 10.1787/44f5453a-en]
27. Izquierdo M, de Souto Barreto P, Arai H, et al. Global consensus on optimal exercise recommendations for enhancing healthy longevity in older adults (ICFSR). J Nutr Health Aging 2025 Jan;29(1):100401. [doi: 10.1016/j.jnha.2024.100401] [Medline: 39743381]
28. Liu BM, Kouladjian O'Donnell L, Redston MR, et al. Association of the Drug Burden Index (DBI) exposure with outcomes: a systematic review. J Am Geriatr Soc 2024 Feb;72(2):589-603. [doi: 10.1111/jgs.18691] [Medline: 38006299]
29. Espinoza SE, Quiben M, Hazuda HP. Distinguishing comorbidity, disability, and frailty. Curr Geriatr Rep 2018 Dec;7(4):201-209. [doi: 10.1007/s13670-018-0254-0] [Medline: 30984516]
30. Kojima G. Frailty as a predictor of disabilities among community-dwelling older people: a systematic review and meta-analysis. Disabil Rehabil 2017 Sep;39(19):1897-1908. [doi: 10.1080/09638288.2016.1212282] [Medline: 27558741]
31. Kojima G, Iliffe S, Walters K. Frailty index as a predictor of mortality: a systematic review and meta-analysis. Age Ageing 2018 Mar 1;47(2):193-200. [doi: 10.1093/ageing/afx162]
32. Roquebert Q, Sicsic J, Santos-Eggimann B, Sirven N, Rapp T. Frailty, sarcopenia and long term care utilization in older populations: a systematic review. J Frailty Aging 2021 Jul;10(3):272-280. [doi: 10.14283/jfa.2021.7]
33. Dapp U, Minder C, Neumann L, Golgert S, Klugmann B, von Renteln-Kruse W. Wirksamkeit der „Aktiven Gesundheitsförderung im Alter". Z Gerontol Geriat 2018 Jun;51(4):379-387. [doi: 10.1007/s00391-018-1392-x]
34. Stuck AE, Moser A, Morf U, et al. Effect of health risk assessment and counselling on health behaviour and survival in older people: a pragmatic randomised trial. PLOS Med 2015 Oct;12(10):e1001889. [doi: 10.1371/journal.pmed.1001889] [Medline: 26479077]
35. Ikonen JN, Eriksson JG, Salonen MK, Kajantie E, Arponen O, Haapanen MJ. The utilization of specialized healthcare services among frail older adults in the Helsinki Birth Cohort Study. Ann Med 2021 Dec;53(1):1875-1884. [doi: 10.1080/07853890.2021.1941232] [Medline: 34714205]
36. Ikonen JN, Eriksson JG, von Bonsdorff MB, Kajantie E, Arponen O, Haapanen MJ. The utilization of primary healthcare services among frail older adults - findings from the Helsinki Birth Cohort Study. BMC Geriatr 2022 Jan 26;22(1):79. [doi: 10.1186/s12877-022-02767-4] [Medline: 35078410]
37. Ilinca S, Calciolari S. The patterns of health care utilization by elderly Europeans: frailty and its implications for health systems. Health Serv Res 2015 Feb;50(1):305-320. [doi: 10.1111/1475-6773.12211] [Medline: 25139146]
38. Roe L, Normand C, Wren MA, Browne J, O'Halloran AM. The impact of frailty on healthcare utilisation in Ireland: evidence from the Irish longitudinal study on ageing. BMC Geriatr 2017 Sep 5;17(1):203. [doi: 10.1186/s12877-017-0579-0] [Medline: 28874140]
39. Roth L, Seematter-Bagnoud L, Le Pogam MA, et al. Identifying common patterns of health services use: a longitudinal study of older Swiss adults' care trajectories. BMC Health Serv Res 2022 Dec 26;22(1):1586. [doi: 10.1186/s12913-022-08987-z] [Medline: 36572888]
40. Sicsic J, Rapp T. Frailty transitions and health care use in Europe. Health Serv Res 2019 Dec;54(6):1305-1315. [doi: 10.1111/1475-6773.13208] [Medline: 31571222]
41. Calciolari S, Luini C. Effects of the bio-psycho-social frailty dimensions on healthcare utilisation among elderly in Europe: a cross-country longitudinal analysis. Soc Sci Med 2023 Dec;339:116352. [doi: 10.1016/j.socscimed.2023.116352]
42. Schnitzer S, Haeger C, Baldenius T, Jürchott K, Behrendt S, Blüher S. Healthcare situation before long-term care - A cohort study of AOK-insured individuals aged 60 years and older in Germany. Gesundheitswesen 2026 Apr 28. [doi: 10.1055/a-2776-2382] [Medline: 42049057]




XSL•FO
RenderX

43. Schulz M, Tsiasioti C, Czwikla J, et al. Claims data analysis of medical specialist utilization among nursing home residents and community-dwelling older people. BMC Health Serv Res 2020 Jul 25;20(1):690. [doi: 10.1186/s12913-020-05548-0] [Medline: 32711516]
44. Wangler J, Jansky M. Disease-Management-Programme: Haben sie die Prävention in der Primärversorgung vorangebracht? Präv Gesundheitsf 2026 Feb;21(1):86-95. [doi: 10.1007/s11553-024-01131-1]
45. Cruz ADO, Santana SMM, Costa CM, Gomes da Costa LV, Ferraz DD. Prevalence of falls in frail elderly users of ambulatory assistive devices: a comparative study. Disabil Rehabil Assist Technol 2020 Jul;15(5):510-514. [doi: 10.1080/17483107.2019.1587016] [Medline: 30907182]
46. Sebban A, Lesclide E, Bonin-Guillaume S, Campana M, Grino M, Franqui C. Previous in-home physiotherapy prevents institutionalization after short-term hospitalization in community-dwelling older dependent people. Aging Clin Exp Res 2020 Jul;32(7):1271-1277. [doi: 10.1007/s40520-019-01310-7] [Medline: 31422566]
47. Nouni-García R, Carbonell-Soliva Á, Orozco-Beltrán D, et al. Association of visiting the physiotherapist with mortality in the spanish general population: a population-based cohort study. Medicina (Kaunas) 2023 Dec 16;59(12):2187. [doi: 10.3390/medicina59122187] [Medline: 38138290]
48. Park JH, Jeon HS, Kim JH, Yoon HB, Lim OB, Jeon M. Immediate effect of insoles on balance in older adults. Foot (Edinb) 2021 Jun;47:101768. [doi: 10.1016/j.foot.2020.101768] [Medline: 33946001]
49. Nickerson KA, Diaz K, Muir BC. Circumstances of falls among older adult walker users in long-term care and the associated walker design deficits. Assist Technol 2024 May 3;36(3):248-253. [doi: 10.1080/10400435.2023.2267623] [Medline: 37796449]
50. Stang A, Jöckel KH. The impact of cancer screening on all-cause mortality. Dtsch Arztebl Int 2018 Jul 23;115(29-30):481-486. [doi: 10.3238/arztebl.2018.0481] [Medline: 30135006]
51. Perkins RB, Wentzensen N, Guido RS, Schiffman M. Cervical cancer screening: a review. JAMA 2023 Aug 8;330(6):547-558. [doi: 10.1001/jama.2023.13174] [Medline: 37552298]
52. Salzman B, Beldowski K, de la Paz A. Cancer screening in older patients. Am Fam Physician 2016 Apr 15;93(8):659-667. [Medline: 27175838]
53. Guittet L, Quipourt V, Aparicio T, et al. Should we screen for colorectal cancer in people aged 75 and over? A systematic review - collaborative work of the French geriatric oncology society (SOFOG) and the French federation of digestive oncology (FFCD). BMC Cancer 2023 Jan 5;23(1):17. [doi: 10.1186/s12885-022-10418-5] [Medline: 36604640]
54. Liss DT, Uchida T, Wilkes CL, Radakrishnan A, Linder JA. General health checks in adult primary care: a review. JAMA 2021 Jun 8;325(22):2294-2306. [doi: 10.1001/jama.2021.6524] [Medline: 34100866]
55. Dapp U, Anders JAM, von Renteln-Kruse W, et al. A randomized trial of effects of health risk appraisal combined with group sessions or home visits on preventive behaviors in older adults. J Gerontol A Biol Sci Med Sci 2011 May;66(5):591-598. [doi: 10.1093/gerona/glr021] [Medline: 21350242]
56. Neumann L, Dapp U, von Renteln-Kruse W, Minder CE. Health promotion and preventive care intervention for older community-dwelling people: long-term effects of a randomised controlled trial (RCT) within the LUCAS cohort. J Nutr Health Aging 2017;21(9):1016-1023. [doi: 10.1007/s12603-017-0932-1] [Medline: 29083443]
57. Conrad A, Valour F, Vanhems P. Burden of influenza in the elderly: a narrative review. Curr Opin Infect Dis 2023;36(4):296-302. [doi: 10.1097/QCO.0000000000000931]
58. Bukhbinder AS, Ling Y, Hasan O, et al. Risk of Alzheimer's disease following influenza vaccination: a claims-based cohort study using propensity score matching. JAD 2022;88(3):1061-1074. [doi: 10.3233/JAD-220361]
59. Zhao H, Zhou X, Fu K, et al. Prospective cohort study evaluating the association between influenza vaccination and neurodegenerative diseases. npj Vaccines 2024;9(1):51. [doi: 10.1038/s41541-024-00841-z]
60. Haeger C, Baldenius T, Schnitzer S, et al. Charakteristika von Antragstellenden bei erstmaliger Feststellung einer Pflegebedürftigkeit – eine bundesweite Analyse von Pflegebegutachtungen von AOK-Versicherten ab 60 Jahren. Z Gerontol Geriat 2025 Feb;58(1):46-54. [doi: 10.1007/s00391-024-02344-x]
61. Castañeda KM, Sidorenkov G, Mourits MJE, et al. Impact of health-related behavioral factors on participation in a cervical cancer screening program: the lifelines population-based cohort. BMC Public Health 2023 Nov 30;23(1):2376. [doi: 10.1186/s12889-023-17293-0] [Medline: 38037016]
62. Lee IC, Chang CS, Du PL. Do healthier lifestyles lead to less utilization of healthcare resources? BMC Health Serv Res 2017 Mar 31;17(1):243. [doi: 10.1186/s12913-017-2185-4] [Medline: 28359277]
63. Kojima T, Hamaya H, Ishii S, Hattori Y, Akishita M. Association of disability level with polypharmacy and potentially inappropriate medication in community dwelling older people. Arch Gerontol Geriatr 2023 Mar;106:104873. [doi: 10.1016/j.archger.2022.104873]
64. Ferreira P, Ferreira AR, Barreto B, Fernandes L. Is there a link between the use of benzodiazepines and related drugs and dementia? A systematic review of reviews. Eur Geriatr Med 2022 Feb;13(1):19-32. [doi: 10.1007/s41999-021-00553-w] [Medline: 34403113]
65. Tanaka T, Akishita M, Kojima T, Son BK, Iijima K. Polypharmacy with potentially inappropriate medications as a risk factor of new onset sarcopenia among community-dwelling Japanese older adults: a 9-year Kashiwa cohort study. BMC Geriatr 2023 Jun 26;23(1):390. [doi: 10.1186/s12877-023-04012-y] [Medline: 37365526]




XSL•FO
RenderX

66. McMaughan DJ, Oloruntoba O, Smith ML. Socioeconomic status and access to healthcare: interrelated drivers for healthy aging. Front Public Health 2020;8:231. [doi: 10.3389/fpubh.2020.00231] [Medline: 32626678]
67. Gordon EH, Hubbard RE. Frailty: understanding the difference between age and ageing. Age Ageing 2022 Aug 2;51(8). [doi: 10.1093/ageing/afac185]
68. Hoffmann F, Icks A. Structural differences between health insurance funds and their impact on health services research: results from the Bertelsmann Health-Care Monitor. Gesundheitswesen 2012 May;74(5):291-297. [doi: 10.1055/s-0031-1275711] [Medline: 21755492]
69. Hoffmann F, Koller D. Different regions, differently insured populations? Socio-demographic and health-related differences between insurance funds. Gesundheitswesen 2017 Jan;79(1):e1-e9. [doi: 10.1055/s-0035-1564074] [Medline: 26492391]

## Abbreviations

**COPD:** chronic obstructive pulmonary disease
**DMP:** disease management program
**GP:** general practitioner
**LTC:** long-term care
**OR:** odds ratio
**PIM:** potentially inappropriate medications
**SHI:** statutory health insurance
**STROBE:** Strengthening of the Reporting of Observational Studies in Epidemiology

*Edited by M Balcarras; submitted 27.Oct.2025; peer-reviewed by B Roy, K Abe, R Lin; revised version received 14.Apr.2026; accepted 14.Apr.2026; published 20.Jul.2026.*

*Please cite as:*
*Baldenius T, Jürchott K, Haeger C, Behrendt S, Blüher S, Schnitzer S*
*Health Care–Related Determinants of First-Time Long-Term Care Need in Older Adults in Germany: Retrospective Cohort Study Using Claims Data*
*Interact J Med Res 2026;15:e86572*
*URL: https://www.i-jmr.org/2026/1/e86572*
*doi:10.2196/86572*

© Till Baldenius, Kathrin Jürchott, Christine Haeger, Susann Behrendt, Stefan Blüher, Susanne Schnitzer. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 20.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Barriers and Facilitators to Physician-Patient Communication in Chinese Tertiary Hospitals From the Perspectives of Hospital-Based Physicians and Patient Relations Coordinators: Qualitative Study

Zengping Shi[1,2,3], PhD; Qinqin Jiang[1,2,4], PhD; Xincheng Wang[2,5], BD; Hongli Yan[1,2], PhD; Yi Xia[6], PhD; Lushaobo Shi[2,4], PhD; Dong Wang[1,2,3,4,5], PhD

[1]School of Public Health, Southern Medical University, No.1023 Shatai Road, Guangzhou, Guangdong, China

[2]School of Health Management, Southern Medical University, Guangzhou, Guangdong, China

[3]The Public Health Policy Research and Evaluation Key Laboratory Project of the Philosophy and Social Sciences of Guangdong College, Southern Medical University, Guangzhou, Guangdong, China

[4]Southern Medical University Center for Health Policy and Governance (Guangdong Provincial Social Science Research Base), Southern Medical University, Guangzhou, Guangdong, China

[5]Research Base for Development of Public Health Service System of Guangzhou, Southern Medical University, Guangzhou, Guangdong, China

[6]Center for Faculty Development and Research, Guangzhou Medical University, Guangzhou, Guangdong, China

**Corresponding Author:**
Dong Wang, PhD
School of Public Health, Southern Medical University, No.1023 Shatai Road, Guangzhou, Guangdong, China

## Abstract

**Background:** Effective physician-patient communication is essential for building trust and sustaining positive relationships, yet becomes increasingly challenging in China's tertiary hospitals, where physicians face heavy workloads.

**Objective:** This study explored the barriers and facilitators of physician-patient communication by synthesizing perspectives from Chinese hospital-based physicians and patient relations coordinators.

**Methods:** A qualitative study using semistructured interviews was conducted with 17 participants (11 physicians and 6 patient relations coordinators) from tertiary hospitals in China (April-July 2025). Data were analyzed using thematic analysis following Braun and Clarke's framework.

**Results:** Barriers and facilitators of physician-patient communication were organized into a multilevel framework comprising patient-level, physician-level, and system-level factors. Patient-level barriers included individual background differences, inadequate expression and limited health literacy, and psycho-emotional states with expectation misalignment, whereas facilitators included effective expectation management, good health literacy and communication readiness, and trust in physicians with shared decision-making. Physician-level barriers involved communication style deficiencies, empathy gaps, and time pressure constraints; facilitators included active listening and patient-centeredness, empathy and emotional support, and clear explanations with cultural adaptability. At the system level, barriers included hospital environment and medical visit settings, legal and policy deficiencies, insufficient communication training, and media-driven distrust with digitally mediated information challenges, while facilitators included institutional support and security assurance, educational training and policy promotion, process optimization and patient health education, and artificial intelligence–assisted baseline knowledge.

**Conclusions:** Physician-patient communication is influenced by multiple factors, necessitating comprehensive intervention measures: enhancing patient education, improving physician communication skills, and strengthening organizational support systems. Notably, special attention should be directed toward addressing the unique challenges posed by digital technologies while concurrently leveraging the opportunities they present to optimize communication outcomes.

***(Interact J Med Res 2026;15:e87947)*** doi:10.2196/87947

**KEYWORDS**

physician-patient communication; communication barriers; communication facilitators; qualitative research; tertiary hospitals



XSL•FO
RenderX

## Introduction

As global health care systems increasingly contend with the challenges of aging populations, constrained resources, and growing patient expectations, effective physician-patient communication has become more crucial than ever, serving as a cornerstone of high-quality health care delivery [1,2]. A growing body of evidence shows that high-quality communication enhances patient satisfaction, improves treatment adherence, reduces medical errors, and ultimately contributes to better health outcomes [3-5]. Conversely, communication failures can have serious consequences; for instance, Humphrey et al [6] reported that an analysis of malpractice claims revealed that 49% involved communication failures and 53% specifically reflected physician-patient miscommunication.

Effective physician-patient communication is inherently a 2-way interaction in which both physicians and patients seek to understand and be understood [7-9]. However, current research on improving physician-patient communication has focused mainly on meeting patients' needs, with particular emphasis on improving physicians' communication skills [10-12]. This patient-centered approach has prioritized patient-reported experiences and satisfaction measures [13] while largely neglecting physicians' reciprocal need to be understood and their perspectives on physician-patient communication challenges.

Moreover, previous studies on physician-patient communication have mostly focused on Western family medical settings, with limited attention to the Chinese context. Yet the determinants of physician-patient communication vary significantly across cultural contexts [14]. Especially within China's inverted triangle health care system [15], tertiary hospitals serve as regional medical centers characterized by high patient volumes, short consultation times, and cross-regional patient mobility. Compared to secondary and primary hospitals, physicians at tertiary hospitals must balance heavy clinical workloads with patients' expectations for detailed explanations [16]. At the same time, the digital transformation of health care, including increased access to online health information, platform-based services, and emerging technologies such as artificial intelligence (AI), is reshaping the context in which physician-patient communication takes place. While these developments may create new opportunities for information exchange, they may also intensify misunderstandings, alter expectations, and introduce new communication challenges that remain insufficiently understood [17,18]. As frontline practitioners experiencing the dual pressures of system strain and digital transformation, physicians in tertiary hospitals may offer valuable insights into these evolving communication challenges. In addition, patient relations coordinators, who are directly involved in managing complaints and mediating disputes, can provide important perspectives on how these challenges manifest in practice and where systemic gaps may exacerbate tensions.

This study explored barriers and facilitators to physician-patient communication from the perspectives of frontline physicians and patient relations coordinators in Chinese tertiary hospitals. By integrating a dual-role perspective, this study aimed to develop a more nuanced understanding of the complex factors shaping physician-patient communication in contemporary hospital settings.

## Methods

### Study Design and Participants

We conducted a qualitative descriptive study using individual in-depth interviews to explore the perspectives of physicians and patient relations coordinators in Chinese tertiary hospitals regarding physician-patient communication. This study was situated within a critical realist ontological framework, which acknowledges that participants' experiences reflect real phenomena in clinical settings, while recognizing that our understanding of these phenomena is inevitably mediated by individual perspectives, social contexts, and the interpretive lens of the research team [19]. The study followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist [20] and was conducted from April to July 2025 in Guangdong Province, China.

Participants were recruited from multiple tertiary hospitals through purposive sampling. Recruitment was initiated by the first authors through direct solicitation, under the guidance of the corresponding author, and was subsequently expanded during data collection to ensure diversity in institutional contexts and participant characteristics. The sampling strategy aimed to achieve variation in professional roles, gender, age, years of experience, and clinical specialty, thereby capturing a broad range of perspectives on physician-patient communication.

A total of 17 participants were recruited from tertiary hospitals, including 11 physicians from 7 tertiary hospitals and 6 patient relations coordinators from 5 tertiary hospitals, representing 11 tertiary hospitals in total across the full sample. The distribution of participants across institutions is detailed in Table 1. These 2 participant groups were purposively selected because they offered distinct yet complementary perspectives on physician-patient communication. Practicing physicians across different clinical departments were recruited to capture firsthand physician-patient communication experiences. Patient relations coordinators were included because of their roles in handling complaints, coordinating communication, mediating disputes, and identifying institutional barriers affecting physician-patient communication. Including both groups enabled the study to integrate direct clinical perspectives with insights from the hospital organizational level. All participants were currently employed, fluent in Mandarin, and had a minimum of 1 year of professional experience to ensure sufficient exposure to physician-patient communication dynamics. Retired practitioners were excluded from the study to ensure perspectives reflected current practice contexts.



XSL•FO
RenderX

**Table .** Sample characteristics (N=17).

| ID | Gender | Age (y) | Years of service | Hospital ID | Department | Education level | Professional title | Administrative position |
|---|---|---|---|---|---|---|---|---|
| Physicians (n=11) | | | | | | | | |
| Dr01 | Male | 46 | 17 | H1 | General Medicine | Master | Associate Senior | Deputy Director of Department |
| Dr02 | Male | 48 | 26 | H1 | Respiratory Medicine | Bachelor | Senior | Deputy Director of Department |
| Dr03 | Female | 35 | 10 | H2 | General Medicine | Bachelor | Junior | None |
| Dr04 | Female | 46 | 14 | H3 | Pulmonary Function Laboratory | Doctor | Associate Senior | None |
| Dr05 | Male | 29 | 3 | H4 | Ophthalmology | Bachelor | Junior | None |
| Dr06 | Male | 42 | 15 | H5 | Traditional Chinese Medicine (Ancient Formulas) | Master | Intermediate | None |
| Dr07 | Male | 50 | 30 | H5 | Cardiology | Doctor | Associate Senior | Director of Department |
| Dr08 | Male | 47 | 22 | H6 | General Medicine | Bachelor | Intermediate | None |
| Dr09 | Male | 38 | 10 | H6 | Cardiology | Master | Intermediate | None |
| Dr10 | Male | 38 | 10 | H7 | Spinal Surgery | Doctor | Associate Senior | None |
| Dr11 | Female | 35 | 10 | H7 | Obstetrics and Gynecology | Bachelor | Intermediate | None |
| Patient relations coordinators (n=6) | | | | | | | | |
| A01 | Female | 31 | 8 | H8 | Physician-Patient Relations Office | Bachelor | None | None |
| A02 | Male | 37 | 16 | H9 | Physician-Patient Relations Office | Bachelor | Intermediate | Deputy Director of Department |
| A03 | Female | 39 | 3 | H10 | Physician-Patient Relations Office | Bachelor | None | None |
| A04 | Female | 33 | 8 | H6 | Physician-Patient Relations Office | Master | None | None |
| A05 | Female | 37 | 16 | H6 | Physician-Patient Relations Office | Bachelor | Junior | None |
| A06 | Female | 27 | 1 | H11 | Physician-Patient Relations Office | Master | None | None |

Recruitment and data collection proceeded iteratively. The study sample size was informed by the concept of information power [21], which considers the study aim, sample specificity, theoretical framework, quality of dialogue, and analytic strategy. Given the focused aim of the study, the use of reflexive thematic analysis, and a purposively diverse sample drawn from 2 relevant participant groups with direct experience of physician-patient communication processes, we anticipated that a relatively modest sample would yield sufficient depth. Throughout data collection, we monitored the richness and



XSL•FO
RenderX

relevance of interview material in relation to the study aims. We concluded recruitment at 17 participants when the research team collectively assessed that the interviews had generated sufficiently rich and diverse data to develop coherent analytic themes addressing the research question, and that additional interviews were unlikely to substantially alter the developing analysis.

## Data Collection and Procedures

Data were collected through semistructured, one-to-one interviews. Participants who could not meet in person were interviewed online via the Tencent Meeting platform, while others participated in face-to-face interviews. All interviews were conducted in Mandarin, lasted 45 - 90 minutes, and were led primarily by ZS, with HY moderating 3 interviews. Demographic data were collected verbally at the beginning of each interview. All interviews were audio- or video-recorded with participants' consent, transcribed verbatim by professional transcriptionists within 24 hours of completion to preserve contextual detail and support timely familiarization with the data. A random sample of 20% of transcripts was independently verified against original recordings by research team members to ensure transcription accuracy. Participant quotes selected for inclusion were translated into English by one bilingual research team member (ZS). To ensure conceptual equivalence and linguistic accuracy, a second independent bilingual team member (QJ) who had no prior exposure to the original Chinese text backtranslated all quotes into Mandarin. Discrepancies between the original and backtranslated versions were reviewed and resolved through consensus discussion among the research team until conceptual equivalence was confirmed. Participants received RMB 150 (approximately US $20) as compensation for their time.

A semistructured interview guide was developed collaboratively by the research team (Multimedia Appendix 1) and refined iteratively based on insights from initial interviews. The interview guide explored four core domains: (1) key physician behaviors and patient preparations that support high-quality interactions, (2) patients' main needs and expectations, (3) characteristics of effective patient-physician communication from the physician's perspective, and (4) personal, situational, and systemic barriers and facilitators to effective physician-patient communication. Throughout the interviews, the research team used techniques such as probing questions, clarification requests, and reflective listening [22] to elicit in-depth responses and encourage participants to share nuanced insights about the facilitators and barriers to physician-patient communication.

To ensure the scientific rigor and quality of data collection, several measures were implemented. All interviewers (ZS and HY) underwent training in qualitative interview techniques and conducted mock interviews prior to data collection to standardize approaches and ensure consistency. All interviews were conducted in quiet, private settings with standardized protocols, including consistent opening statements and informed consent procedures. For online interviews, technical equipment was tested beforehand to ensure optimal audio and video quality. Interviewers maintained brief reflective notes during and immediately after each interview to capture contextual observations, nonverbal cues where available, and preliminary reflexive insights.

## Researcher Reflexivity Statement

Given the interpretive nature of reflexive thematic analysis, it is important to acknowledge the positionality and perspectives that the research team brought to this study [23,24]. The lead researcher and primary interviewer (ZS) is a doctoral researcher with training in qualitative research methods and prior professional experience in culture- and communication-related work at a tertiary hospital in China. This background provided familiarity with the dynamics of physician-patient communication in China's health care system, which facilitated rapport-building with participants and sensitivity to contextual nuances during interviews. At the same time, this proximity required reflexive awareness of potential assumptions about what constitutes "effective" communication, and ZS took care to remain open to perspectives that might differ from her own professional experiences.

The broader research team comprised members with expertise in clinical medicine, health communication, qualitative methodology, and health services research, which enriched the analytic process by bringing multiple interpretive lenses to the data. ZS maintained brief reflexive notes after each interview and met regularly with the supervisory team to discuss emerging interpretations, challenge assumptions, and consider how the team's positionality may have shaped the developing analysis. These reflexive practices were integral to the analytic process and supported the generation of themes that were grounded in participants' accounts while acknowledging the research team's active role in interpretation [19].

## Data Analysis

Interview transcripts were analyzed using reflexive thematic analysis following Braun and Clarke's 6-step framework [23-25], which emphasizes the researcher's active and reflexive role in generating themes through deep engagement with the data. An inductive, data-driven orientation was adopted, whereby codes and themes were developed in response to participants' accounts without the imposition of a predetermined framework. Data management and coding were conducted using NVivo 20 (Lumivero LLC). Participant quotes were coded using alphanumeric identifiers: "Dr" followed by numbers (Dr01-Dr11) for practicing physicians, and "A" followed by numbers (A01-A06) for members from physician-patient relations management departments.

ZS led the analysis, conducting inductive coding of all transcripts through an iterative process of close reading, annotation, and code development. To deepen analytic reflexivity and broaden interpretive engagement with the data, XW independently coded a subset of transcripts (the first 2 transcripts from each participant group). These additional coding contributions informed collaborative discussions about alternative readings and emerging patterns, contributing to the refinement of the developing analytic framework. The purpose of this collaborative coding was not to achieve consensus or establish interrater agreement, but rather to bring multiple



XSL•FO
RenderX

interpretive perspectives to the data and to deepen the team's reflexive engagement with the analysis.

The team met regularly during the analytic process to discuss coding decisions, review evolving themes, and critically examine the developing analysis. DW provided an additional interpretive perspective during theme refinement. The final thematic framework was developed through an iterative process of deep engagement with the data across all phases of analysis, moving between data, codes, and candidate themes in a recursive manner. Themes were generated as interpretive patterns of shared meaning across the dataset, rather than predetermined categories [23]. In keeping with reflexive thematic analysis, analytic rigor was supported through sustained engagement with the full dataset, reflexive memoing, and regular interpretive discussions within the research team, rather than through coder agreement or the resolution of coding discrepancies to achieve consensus.

### Trustworthiness and Quality

Several strategies were used to enhance the trustworthiness of the study, guided by considerations of credibility, transferability, dependability, and confirmability [26], while remaining attentive to the reflexive and interpretive nature of the analytic approach [23].

Credibility was supported through prolonged engagement with the data, purposive sampling to capture diverse perspectives across clinical specialties and professional roles, and the inclusion of 2 complementary participant groups (physicians and patient relations coordinators) to provide a multifaceted understanding of physician-patient communication. Collaborative coding of a subset of transcripts and regular team discussions further enriched the analytic process by introducing multiple interpretive perspectives.

Transferability was enhanced through a detailed description of the study context, participant characteristics, and analytic procedures, enabling readers to assess the applicability of the findings to their own settings. The purposive sampling strategy ensured diversity in institutional contexts, clinical specialties, and professional roles across the sampled tertiary hospitals in Guangzhou and Shenzhen, Guangdong Province, China.

Dependability was addressed through transparent documentation of all research decisions throughout the study. The interview guide, coding process, and thematic development were documented and discussed within the research team, creating an audit trail that traces the analytic journey from raw data to final themes.

Confirmability was supported through reflexive practices, including the maintenance of reflexive notes by interviewers, regular supervisory discussions about emerging interpretations, and the explicit acknowledgment of the research team's positionality and its potential influence on the analysis. Within the framework of reflexive thematic analysis, we recognize that analysis is inherently shaped by the researcher's subjectivity; therefore, confirmability was pursued not through the elimination of researcher influence, but through transparency about the interpretive process [19,27].

### Ethical Considerations

This study was approved by the Ethics Committee of Southern Medical University, Guangzhou, Guangdong Province, China (approval number NFYKDX003-2025-35), and was conducted in accordance with the Declaration of Helsinki. Prior to the interviews, participants were informed about the purpose of the study and the use of audio recording, and informed consent was obtained from all participants. Interviews were audio-recorded only with participants' permission. No personally identifiable information was collected or reported, and all data were kept confidential.

## Results

### Demographics of the Participants

We interviewed a total of 17 participants, comprising 11 physicians and 6 patient relations coordinators (Table 1). Among the physicians, there were 8 males and 3 females, with a mean age of 41.27 (SD 6.40; range 29‐50) years and a mean professional experience of 15.18 (SD 7.65; range 3‐30) years. They represented a broad range of specialties, including obstetrics and gynecology, cardiology, spinal surgery, general medicine, and other specialties. Given China's high patient-to-physician ratio and the central role of tertiary hospitals in delivering specialist care, physicians with senior roles (eg, department directors or deputy directors) remain active frontline clinicians regardless of administrative responsibilities. Three physicians with administrative roles reported conducting outpatient consultations on a regular basis, averaging 3 to 4 sessions per week. Among the patient relations coordinators, there were 1 male and 5 females, with a mean age of 34.00 (SD 4.12; range 27‐39) years and a mean professional experience of 8.67 (SD 5.61; range 1‐16) years.

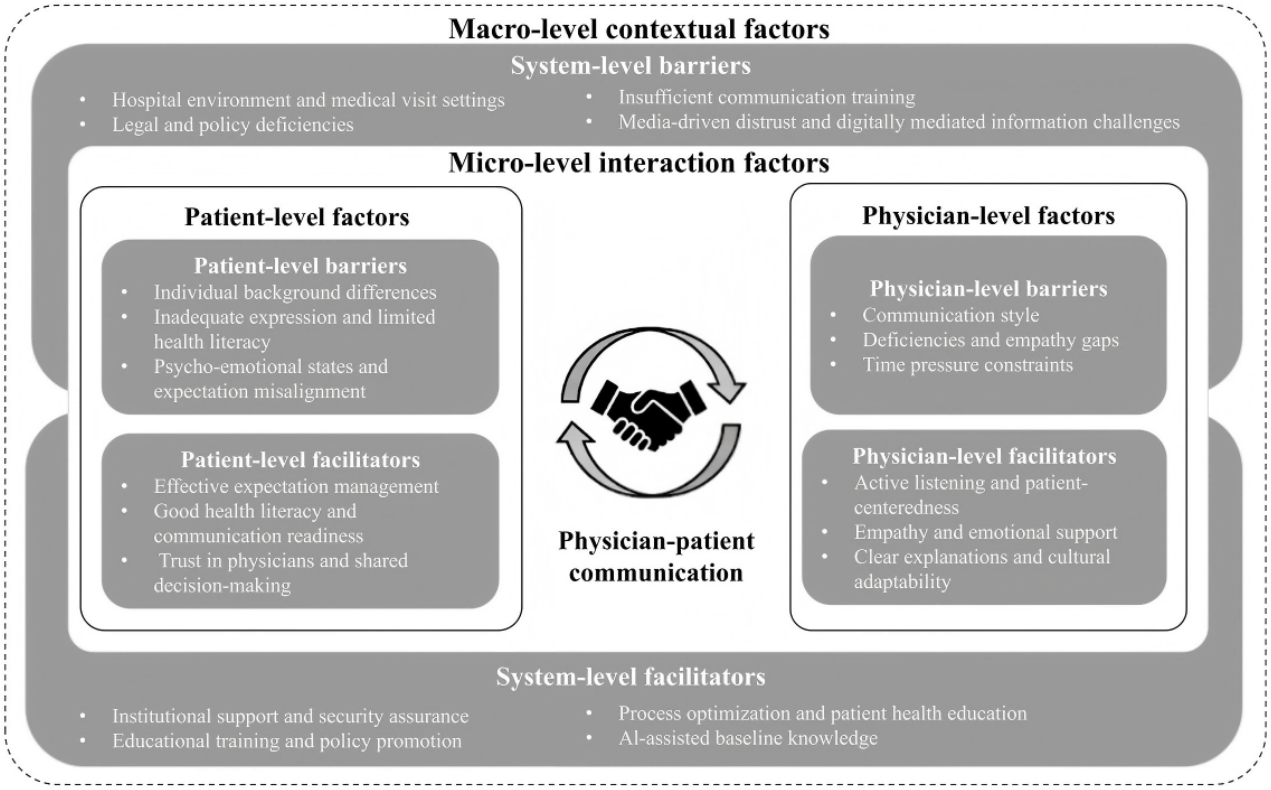
### Overview of Themes

Thematic analysis of the interview data identified 2 overarching themes: barriers and facilitators of effective physician-patient communication. As shown in Figure 1, these themes were organized into a multilevel framework comprising patient-level, physician-level, and system-level factors. Patient- and physician-level factors directly shaped the communication process, whereas system-level factors functioned as the broader context that could either hinder or enable effective interaction. The following sections discuss each theme and its constituent subthemes in detail.



XSL•FO
RenderX

**Figure 1.** Overview of the themes. AI: artificial intelligence.




## Theme 1: Barriers to Effective Physician-Patient Communication

### *Patient-Level Barriers*

#### Individual Background Differences

Physicians reported that patients' educational levels directly influenced communication effectiveness, with particularly notable differences between highly educated and less educated patients. Many highly educated patients tended to arrive with strong preformed views about their conditions, sometimes displaying overconfidence and questioning physicians' clinical judgments. Conversely, patients with limited education required extensive, repeated explanations to understand medical information, frequently resulting in miscommunication. For these patient groups, the standardized communication methods taught in medical schools were largely ineffective, and time constraints in high-volume outpatient settings made it difficult to tailor communication strategies to individual patients. As one physician explained,

> *For some well-educated patients, the problem lies in overconfidence; for others, I need to explain and rephrase multiple times before they can understand what I mean.* [Dr05]

Patient age also shaped communication dynamics. Younger patients demonstrated higher autonomous learning motivation and were more likely to form preliminary self-diagnoses before consultations, a tendency exacerbated by the digital information environment elaborated in the subsequent section (see "System-Level Barriers"). Furthermore, they often approached health care from a consumerist perspective, expecting guaranteed outcomes and engaging in what participants described as excessive rights protection behaviors, such as recording consultations and filing complaints. Such behaviors often prompted physicians to adopt defensive communication practices, including avoiding discussions of clinical uncertainty. As a member of the physician-patient relations office noted,

> *Many young patients treat medical consultations like 'online shopping,' expecting guaranteed treatment outcomes and a 'seven-day free return' policy. However, medicine is inherently fraught with uncertainty, and no one can guarantee outcomes...Some patients record every interaction, forcing physicians to speak with excessive caution and hesitate to discuss treatment uncertainties.* [A03]

#### Inadequate Expression and Limited Health Literacy

Inadequate symptom expression and limited health literacy are the primary patient-level barriers to physician-patient communication. Physicians reported that many patients lacked the ability to accurately describe symptoms and express their core consultation needs clearly. Patients often used colloquial language, requiring physicians to adjust their questioning strategies repeatedly and clarify complaints through follow-up inquiries. Even so, patients frequently provided irrelevant responses, resulting in the loss of key clinical information. One physician elaborated on the difficulty of obtaining effective clinical information:

> *When I ask under what specific circumstances the stomach pain occurs, they just repeat 'severe bloating' without answering the actual question...Sometimes we spend five minutes and still cannot obtain the necessary information, which leaves us feeling quite helpless.* [Dr07]



XSL•FO
RenderX

One physician further noted,

> *Patients who can clearly express their conditions and are quite professional are actually rare... The health literacy of most patients in China still has significant room for improvement.* [Dr01]

Beyond incomplete symptom descriptions, participants emphasized a significant disconnect between patients' surface-level expressions and their underlying needs. One physician added,

> *Some patients are unable to clearly articulate their true needs; in fact, what a patient says during a consultation may not reflect their actual needs. If the doctor cannot discern this, communication will still fail even if the diagnosis is correct.* [Dr05]

#### Psycho-Emotional States and Expectation Misalignment

Physicians described encountering contradictory patient demands and misaligned expectations during treatment processes. On one hand, patients' psycho-emotional status, such as disease-related anxiety, significantly disrupted communication. One physician provided a vivid example:

> *A patient might just have a sore throat, but it is accompanied by overwhelming anxiety, fearing they would be reprimanded by their boss, and might even lose their job.* [Dr03]

Such intense emotional states caused patients to talk excessively and pose repetitive questions, overwhelming the physician and impeding effective clinical communication. On the other hand, expectation misalignment was also evident when patients sought definitive diagnoses or guaranteed outcomes while resisting the examinations or costs required to support clinical decision-making. As one physician stated,

> *Some patients hope for both definitive diagnosis and cost savings, which is somewhat contradictory in certain situations. For example, to give a definitive diagnosis, I need to order some examinations to find the root cause. Naturally, this requires more examinations, but once examinations increase, patients feel that costs might be higher, so contradictions exist here.* [Dr01]

### Physician-Level Barriers

#### *Communication Style Deficiencies and Empathy Gaps*

Physicians' professional capabilities directly influenced their ability to understand and manage patient conditions. Good professional competence helped enhance patient trust and treatment compliance; conversely, insufficient professional capabilities not only weakened therapeutic effects but also reduced communication effectiveness. As one physician noted,

> *If I cannot clearly explain or show confidence in diagnosis, patients will naturally doubt what I say.* [Dr02]

Some physicians exhibited authoritative tendencies in communication, often assuming patients could understand medical concepts while lacking the ability to translate professional information into accessible language. Communication involved excessive use of medical terminology and rigid language, tending toward unidirectional information delivery rather than shared decision-making with patients, sometimes displaying condescending attitudes. As a member of the physician-patient relations office observed,

> *Many physicians tend to dominate conversations. They use many professional terms, but patients often have no idea what they mean.* [A06]

This approach not only weakened patient participation but also exacerbated power imbalances between physicians and patients, making patients feel ignored or disrespected. Furthermore, when facing physician-patient conflicts, some physicians lacked awareness of appropriate vulnerability demonstration and empathy expression, requiring improvement in communication skills. A physician remarked,

> *Physician-patient conflicts often escalate from verbal confrontations. In some cases, some conflict escalations occur when physicians insist on verbal dominance and want to win the argument. When patients feel disadvantaged in communication process, they may turn to other forms of confrontation, such as calling 12,315 (Chinese consumer rights complaint hotline), or engaging in physical confrontation. In my view, physicians do not need to compete for victory; instead, they can show appropriate vulnerability and defuse conflicts before they escalate beyond verbal confrontation.* [Dr02]

#### *Time Pressure Constraints*

Insufficient time was the most frequently mentioned communication barrier by physicians. Participants' accounts showed that high patient volume, limited staffing, and heavy workload reduced the time physicians could allocate to each consultation. Several physicians described that, although they valued communication and recognized its importance in clinical care, time pressure often constrained them to provide detailed condition explanations. For instance, one physician stated,

> *Patients might wonder why physicians can't give them more time. I also want to give them plenty of time, but I need to see 50 patients in one morning, with over ten people still waiting. My work pressure is also enormous; I can only try my best to do well.* [Dr05]

Likewise, one physician noted,

> *it's almost time to clock out, and there are still 10 to 20 people are queuing behind you...objectively, there is simply no time.* [Dr09]

This perception was also echoed by a member of the physician-patient relations office, who commented,

> *Some physicians see 50 patients in a single morning. Do they realistically have the time to explain everything to each patient individually? Not possible.* [A06]



XSL•FO
RenderX

## System-Level Barriers

### *Hospital Environment and Medical Visit Settings*

Any problems in the medical care-seeking process, such as long waiting times or cumbersome examinations leading to poor visit experiences, could cause patients to project emotions onto physicians with whom they directly interacted. Emotional venting was one of the important causes of physician-patient communication conflicts. Some patients carried emotions from negative previsit experiences into communications with physicians. A member of the physician-patient relations office described,

> *The tipping point for a patient's emotional outburst is rarely an isolated incident; rather, it is the culmination of their preceding experiences. For example, patients waited one or two hours to see the physician, the consultation is over in just two or three minutes. After that, patients have to go through examinations, wait for reports, make payments, and return visits. The whole process is very exhausting.* [A02]

### *Insufficient Communication Training*

Although medical education invested heavily in clinical skill development, physician-patient communication training remained relatively underdeveloped, with curriculum content primarily focused on diagnosis and technical operations, while communication training for high-complexity clinical scenarios and patient emotional and psychological needs remained insufficient. This problem was particularly prominent among young physicians. Limited by clinical experience, they often lacked confidence in communication and struggled to quickly establish trust; simultaneously, patients questioned the professional authority of junior physicians, further exacerbating communication barriers. A member of the physician-patient relations office noted,

> *In school, physician-patient communication was not a core curriculum. Although hospitals provide training after employment, genuine communication skill improvement still depends on personal understanding. Because communication scenarios are complex, merely teaching communication theory cannot actually solve practical problems. Communication is a capability requiring lifelong learning.* [A04]

### *Legal and Policy Deficiencies*

Insufficient legal enforcement and protection mechanisms left both patients and physicians feeling vulnerable. Physicians reported that patients currently possessed numerous accessible channels for expressing demands; however, physicians lacked adequate advocacy channels when facing unreasonable demands, verbal abuse, or violence. This "legal vacuum" forced medical staff to adopt defensive postures, hindering honest dialogue about medical uncertainty or treatment limitations. Additionally, complex medical insurance policies (such as reimbursement scope and self-payment ratios) often became "noise" in physician-patient communication. Prescription authority restrictions, medical insurance reimbursement clauses, or hospital regulations often forced physicians to refuse patient requests. This "institutional rejection" was easily misunderstood by patients as physician indifference, exacerbating relationship tensions. As one physician stated,

> *If I cannot be certain of receiving protection when problems arise, I will avoid discussing any potentially controversial content.* [Dr11]

One physician expressed similar difficulties,

> *Sometimes I must say 'no' due to regulations, but patients think I simply don't care.* [Dr04]

### *Media-Driven Distrust and Digitally Mediated Information Challenges*

Participants described media narratives and digital health information environments as important barriers to effective physician-patient communication. For one thing, biased media coverage of negative medical incidents was found to foster patients' preconceived distrust of physicians prior to consultations, which directly hindered the establishment of therapeutic rapport, mutual trust, and open communication. As one physician observed,

> *Some media's negative medical reporting affects patients' impressions of physicians. Patients believe physicians are making money from patients' illnesses, arriving with such preconceived notions before visits. In such situations, it's very difficult for physicians to change these prejudices through communication, so public opinion direction is extremely important.* [Dr07]

On the other hand, the growing use of digital technologies and the wide availability of online health information introduced new communication challenges. Physicians reported that patients increasingly searched for health information online and, in some cases, used AI tools such as DeepSeek (Hangzhou DeepSeek Artificial Intelligence Co, Ltd) before or after consultations. Patients were often described as becoming partially informed through fragmented online or AI-generated information, without fully understanding the broader clinical context. This increased the explanatory burden on physicians and sometimes shifted consultations away from treatment planning toward the correction of misinformation. As one physician explained,

> *More and more patients come with DeepSeek recommendations. Although I'm willing to maintain an open attitude to communication, each time I need to spend time mutually sharing materials and judging which side is more reasonable and has stronger evidence, it's quite exhausting. Sometimes I spend more time explaining why AI recommendations aren't suitable for patients than discussing actual treatment plans.* [Dr09]



XSL•FO
RenderX

## Theme 2: Facilitators of Effective Physician-Patient Communication

### *Patient-Level Facilitators*

#### Effective Expectation Management

Effective management of expectations was identified as a key characteristic of patients that facilitates constructive physician-patient communication. Participating physicians clearly preferred patients who understood the limitations of medicine, the complexities of diagnosis, and the uncertainties of treatment outcomes. As a member of the physician-patient relations office explained,

> *Many patients expect health to be a completely disease-free state. But medicine has its limitations. Physicians can only promise to improve conditions through treatment; a complete absence of disease is impossible. If patients can manage their expectations, we can have peaceful discussions about what is possible, rather than being forced to make promises that we cannot fulfill.* [A01]

Physicians noted that such patients enabled more open and honest discussions about treatment options, potential outcomes, and associated risks.

#### Good Health Literacy and Communication Readiness

Patients' strong communication skills and adequate health literacy helped improve physician-patient interaction quality. Patients who could clearly present medical histories, ask targeted questions, and demonstrate understanding of basic medical concepts enabled physicians to focus more on complex decisions and personalized treatment.

> *If patients can clearly describe symptoms and understand basic terminology, we can spend more time developing treatment plans rather than starting from scratch.* [Dr11]

Patients' strong sense of health responsibility was also emphasized as a valuable factor promoting physician-patient communication. Physicians valued well-prepared patients who maintained organized medical records, obtained information from reliable sources, and asked clear questions.

> *A patient who communicates well knows they are the primary person responsible for their own health, rather than believing that medical visits are entirely the physician's responsibility. Patients can prepare their past medical history materials as much as possible, allowing us to communicate and interact, examine disease development trends, making the entire diagnostic and treatment process smoother and more efficient.* [Dr09]

#### Trust in Physicians and Shared Decision-Making

Patient trust in physicians and shared decision-making more easily helped patients and physicians establish balanced and mutually trusting dialogue atmospheres. Physicians emphasized preferring patients who respected their professional knowledge while asking insightful questions, with shared decision-making during communication interactions being more aligned with patient needs.

> *Trust is the foundation. When patients trust physicians but still ask questions, this means they are actively participating in communication. This type of communication interaction more easily facilitates shared decision-making, helping patients choose treatment plans most suitable for their current personal situations.* [Dr01]

### Physician-Level Facilitators

### *Active Listening and Patient-Centeredness*

Participating physicians emphasized that active listening and patient-centeredness were foundations of effective medical communication. Through attentive listening, asking follow-up questions based on patients' verbal or nonverbal cues, and summarizing their core concerns, physicians not only made patients feel heard and valued but also reduced risks of missing important symptoms or worries. One physician mentioned,

> *I believe treating illness requires standing from the patient's perspective, first treating the heart. Many patients don't have physical illness but anxiety. When physicians patiently listen to their complete expressions, patients may feel half-better upon leaving the consultation room.* [Dr10]

### *Empathy and Emotional Support*

Empathy and emotional support were repeatedly mentioned as core elements for building trust and enhancing treatment compliance. Physicians who could recognize patient fears, acknowledge their concerns, and provide comfort could establish relationships transcending clinical interventions. Such connections were viewed as key to alleviating anxiety and enhancing patient confidence in treatment processes.

> *I believe a very important aspect of being a doctor is having understanding and empathy, being able to comprehend why patients say what they do; additionally, being able to make patients feel 'this physician understands me' from the first words spoken.* [Dr05]

### *Clear Explanations and Cultural Adaptability*

Using accessible language for clear explanations was viewed as key to bridging gaps between medical expertise and patient understanding. Physicians described the importance of translating professional terminology into colloquial language, using analogies or visual aids when appropriate, and confirming patient comprehension levels to improve communication. One physician emphasized,

> [Many medical professional terms are incomprehensible to patients; we need to translate professional terms into common people's language, making them understandable and clear. Dr02]

Maintaining transparency when discussing diagnoses, treatment options, and prognoses was also viewed as a hallmark of effective communication. Physicians who frankly discussed uncertainties, limitations, and potential complications could



XSL•FO
RenderX

maintain hope while setting realistic expectations, thereby avoiding disappointment and conflict. Many associated this transparency with more collaborative relationships, especially when combined with shared decision-making approaches that incorporated patient preferences and values.

> *I told them frankly what we knew, what we didn't know, and what the risks were, and then we decided together how to proceed.* [Dr11]

When interacting with diverse patient populations, cultural sensitivity and adaptive communication styles were viewed as crucial. Recognizing differences in health beliefs, communication preferences, and decision-making processes enabled physicians to adjust communication approaches accordingly, reducing misunderstandings and promoting respectful, culturally appropriate care. As one physician explained,

> *If you don't understand the patient's cultural background, you may be talking, but you're not really connecting.* [Dr11]

## System-Level Facilitators

### *Institutional Support and Security Assurance*

Physicians unanimously emphasized that institutional support and security assurance provided by medical institutions were important prerequisites for physicians to conduct effective communication. For example, hospitals' zero-tolerance policies toward violence and abuse, along with clear behavioral standards and response mechanisms, made physicians feel more secure and confident when communicating with patients, thereby reducing defensive behaviors and promoting open dialogue. Transparent and efficient complaint handling processes were highly valued for protecting physicians from unfounded accusations while ensuring reasonable patient grievances were addressed. Multidepartmental collaboration and information-sharing mechanisms, along with third-party coordinator intervention (such as physician-patient relations offices, AI assistants, and patient companions), could effectively reduce information gaps and misunderstandings, ensuring continuity and smoothness in diagnostic and treatment processes. As a member of the physician-patient relations office said,

> *Patients always feel that they are the weaker party, but in fact they already have multiple channels for providing feedback. Healthcare workers are quite difficult because various parties may make all kinds of demands on them. Medical institutions should provide health care workers with more psychological, financial, and other forms of support, and ensure their safety, so that healthcare workers can focus on clinical work and form a virtuous cycle.* [A03]

### *Process Optimization and Patient Health Education*

Physicians believed that streamlined processes could help shorten patient waiting times and reduce information delays. Additionally, providing patient health education to empower them to play an active role in their treatment was crucial for maintaining high-quality communication. In particular, patient education and health literacy improvement programs could help patients better understand medical information and actively participate in medical discussions, thereby reducing communication barriers caused by information asymmetry. As one physician explained,

> *If the appointment and check-in processes are straightforward and patients receive basic guidance beforehand, they will arrive less stressed and better prepared. We can use that time for what truly matters, such as collaboratively determining the best next steps.* [Dr11]

### *Educational Training and Policy Promotion*

Participants emphasized that education and policy interventions were key to maintaining and strengthening these facilitating factors. Embedding scenario-based communication training throughout medical education and professional development (such as role-playing, standardized patient simulations, and video reviews) could significantly enhance physicians' practical communication abilities. Mentorship and exemplary demonstrations helped continuously reinforce communication skills in clinical environments, forming positive transmission. At the policy level, through patient rights and responsibility education, health care worker protection legislation, and performance assessments incorporating communication quality indicators, effective communication could become an important measure of medical services, incentivizing physicians to continuously improve communication abilities and ensuring sincere exchanges in safe environments. A member of the physician-patient relations office explained,

> *Physician-patient communication is not just one-time learning, but requires reminders, feedback, and support at every stage of professional careers.* [A05]

### *AI-Assisted Baseline Knowledge*

In contrast to the communication barriers posed by fragmented internet information (as displayed in Theme 1), some physicians hold positive attitudes toward AI-assisted diagnosis when used by patients with adequate health literacy. They viewed it as a superior alternative to traditional, unreliable information sources. Highlighting this nuanced perspective, one physician stated,

> *I generally welcome these learning-oriented patients; at least most AI information is accurate. I'd rather they consult AI than turn to Baidu platform or listen to neighbors.* [Dr08]

Simultaneously, some physicians emphasized that AI diagnosis required professional oversight to reduce risks and ensure patient benefits.

> *I maintain an overall positive attitude toward AI diagnosis. Relying solely on AI for diagnostic decisions carries risks, but these risks can be managed through professional oversight, ultimately benefiting patients at the patient level.* [Dr02]

This divergence illustrates that the digital information environment acts as a double-edged sword, where its impact heavily depends on the patient's underlying health literacy and the physician's guiding approach.



XSL•FO
RenderX

## Discussion

### Principal Findings

This qualitative study explored the barriers and facilitators of physician-patient communication in Chinese tertiary hospitals from dual perspectives of physicians and patient relations coordinators, and organized these factors into a multilevel framework covering patient, physician, and system levels. At the patient level, barriers included individual background differences, inadequate expression and limited health literacy, and psycho-emotional states with expectation misalignment; facilitators included effective expectation management, good health literacy and communication readiness, and trust in physicians with shared decision-making. At the physician level, barriers involved communication style deficiencies and empathy gaps as well as time pressure constraints; facilitators included active listening and patient-centeredness, empathy and emotional support, and clear explanations with cultural adaptability. At the system level, barriers encompassed hospital environment and medical visit settings, legal and policy deficiencies, insufficient communication training, and media-driven distrust with digitally mediated information challenges; facilitators included institutional support and security assurance, educational training and policy promotion, process optimization and patient health education, and AI-assisted baseline knowledge.

Consistent with the propositions of information ecology theory [28], our findings emphasize that physician-patient communication is not determined by individual factors alone but is embedded in contextual environments. This suggests that improving health care interactions cannot be achieved solely by enhancing physicians' professional skills. Similar to previous research [29-31], we found that excessive administrative burdens and conflict management responsibilities undermine physicians' ability to engage in effective communication with patients. Redistributing nonmedical responsibilities to appropriate administrative personnel or digital systems may alleviate physicians' workload, reduce communication fatigue, and ultimately foster more meaningful, patient-centered interactions. Administrators from patient relations coordinators emphasized systemic solutions, including third-party mediation mechanisms, cross-departmental coordination, and a zero-tolerance policy for violent behavior. These measures are often overlooked but are important drivers for improving communication between physicians and patients. These perspectives challenge the dominant focus on individual-level communication training. However, it allows physicians to focus on their core medical responsibilities with peace of mind.

Furthermore, our findings highlight the need to attend to patient-level barriers and facilitators that influence physician-patient communication. As proposed by Inui and Carter [32], physician-patient communication is a reciprocal exchange that demands active participation from both parties. While patient-centered communication requires balanced information exchange and complementary roles, the communication gap often extends beyond simple information asymmetry to include disparities in health literacy, shared understanding, and collaborative care capacity [33,34]. To bridge this gap, comprehensive patient education programs are essential and should address 2 key dimensions. First, education must narrow the knowledge divide between medical professionals and the general public. This entails moving beyond traditional paternalistic models toward a collaborative framework in which patients become informed partners, equipped with basic health literacy, a clear understanding of treatment options, and realistic expectations regarding therapeutic outcomes [35]. Second, such programs should build communication competencies, empowering patients to make the most of limited consultation time by articulating symptoms clearly, asking relevant questions, expressing concerns appropriately, and participating in shared decision-making [36]. When patients proactively express their expectations and concerns, physician-patient communication becomes more efficient and meaningful. This, in turn, fosters mutual tolerance, understanding, and respect, thereby promoting a supportive and reciprocal communicative dynamic [37].

Within a multilevel framework, AI functions as a digitally mediated environmental factor that introduces new challenges to physician-patient communication. Although existing studies have often compared patient preferences for AI- versus physician-generated communication [38], or evaluated the relative quality of AI-assisted and clinician communication [39], less is known about how physicians respond when patients bring AI-generated information into real-world consultations. Our findings address this gap by showing that physicians did not view AI-informed communication as uniformly beneficial or harmful, but evaluated it according to how patients used, interpreted, and communicated AI-generated information during clinical encounters. Consistent with recent scoping review evidence [39], our findings suggest that AI-informed communication was not perceived as uniformly harmful or beneficial. On the one hand, AI may improve patients' baseline knowledge and support more informed questioning, thereby creating conditions for more patient-centered communication. On the other hand, it may also introduce new communication pressures when AI-generated information is inaccurate, decontextualized, or treated by patients as a substitute for clinical judgment. Importantly, physicians in our study did not uniformly reject AI-informed patients. Rather, they distinguished between patients who used AI as a learning tool to better understand their condition and those who used AI-generated advice to challenge professional judgment. This finding suggests that the impact of AI on physician-patient communication depends less on the presence of AI itself than on how AI-generated information is interpreted, communicated, and managed within clinical encounters.

### Strengths and Limitations

Grounded in the context of Chinese tertiary hospitals, this study breaks the limitation of single-stakeholder views by integrating the experiences of practicing physicians with the insights of patient relations coordinators. It bridges micro-level clinical encounter practices with macro-level organizational processes and further identifies the double-edged sword effect of digital transformation on physician-patient communication. Collectively, this work provides constructive reflections for



XSL•FO
RenderX

optimizing physician-patient interactions in high-workload clinical environments.

Several limitations should be noted. First, this study focused on tertiary hospitals in Guangzhou and Shenzhen, 2 economically developed cities in southern China. While these settings provide valuable insights into high-complexity health care environments, findings may not fully reflect communication dynamics in different hospital levels or geographic regions. Second, the qualitative design offers rich contextual understanding but cannot quantify the relative importance of identified factors or establish causal relationships. Third, the study was conducted during a specific timeframe and may not capture evolving digital health trends. Future research should examine communication patterns across diverse health care settings and incorporate mixed methods approaches to validate these findings and assess intervention effectiveness.

## Conclusion

This qualitative study of frontline physicians and patient relations coordinators within tertiary hospitals in China identified the key barriers and facilitators to physician-patient communication. Our findings support that effective physician-patient communication is influenced by a variety of factors, including patient-, physician-, and system-level factors. In the modern era, physician-patient communication is challenged by consumerist medical attitudes, the internet age, and societal changes, but also benefits from new mechanisms such as third-party coordination and technological assistance. To achieve effective communication, it is necessary to understand these complex factors and establish a relationship based on mutual understanding and respect, while also implementing systematic education and training, policy support, and organizational improvements to optimize the communication environment.

## Acknowledgments

The authors would like to express their sincere gratitude to all participants in this study. The authors declare the use of generative artificial intelligence (GenAI) in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: proofreading and editing, and reformatting sentence structure in specific, limited portions of text. The GenAI tool used was ChatGPT-5.4. Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

## Funding

This research was supported by Basic and Applied Basic Research Foundation of Guangdong Province (grant number 2023A1515011124). The funder had no role in the study design; data collection, management, analysis, or interpretation; manuscript preparation, review, or approval; or the decision to submit the manuscript for publication.

## Data Availability

The datasets generated and/or analyzed during the current study are not publicly available to safeguard participant confidentiality but are available from the corresponding author on reasonable request.

## Authors' Contributions

Conceptualization: ZS, QJ, DW
Data curation: ZS, QJ
Formal analysis: ZS, XW
Funding acquisition: DW
Investigation: ZS, HY
Methodology: ZS, QJ, DW
Project administration: ZS, DW
Resources: ZS, DW
Software: ZS, XW
Supervision: ZS, QJ, DW
Validation: ZS, QJ, XW, DW
Visualization: ZS, XW
Writing – original draft: ZS
Writing – review & editing: ZS, QJ, HY, XW, YX, LS, DW

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File, 13 KB - i-jmr_v15i1e87947_app1.docx ]



XSL•FO
RenderX

## References


1. Jones CH, Dolsten M. Healthcare on the brink: navigating the challenges of an aging society in the United States. NPJ Aging 2024 Apr 6;10(1):22. [doi: 10.1038/s41514-024-00148-2] [Medline: 38582901]
2. Keshavarzi MH, Safaie S, Faghihi SAA, Zare S. Barriers of physician-patient relationships in professionalism: a qualitative study. J Adv Med Educ Prof 2022 Jul;10(3):199-206. [doi: 10.30476/JAMP.2022.94010.1563] [Medline: 35910520]
3. Sharkiya SH. Quality communication can improve patient-centred health outcomes among older patients: a rapid review. BMC Health Serv Res 2023 Aug 22;23(1):886. [doi: 10.1186/s12913-023-09869-8] [Medline: 37608376]
4. Cuffy C, Hagiwara N, Vrana S, McInnes BT. Measuring the quality of patient-physician communication. J Biomed Inform 2020 Dec;112:103589. [doi: 10.1016/j.jbi.2020.103589] [Medline: 33035705]
5. Świątoniowska-Lonc N, Polański J, Tański W, Jankowska-Polańska B. Impact of satisfaction with physician-patient communication on self-care and adherence in patients with hypertension: cross-sectional study. BMC Health Serv Res 2020 Nov 16;20(1):1046. [doi: 10.1186/s12913-020-05912-0] [Medline: 33198739]
6. Humphrey KE, Sundberg M, Milliren CE, Graham DA, Landrigan CP. Frequency and nature of communication and handoff failures in medical malpractice claims. J Patient Saf 2022 Mar 1;18(2):130-137. [doi: 10.1097/PTS.0000000000000937] [Medline: 35188927]
7. van Vliet LM, Epstein AS. Current state of the art and science of patient-clinician communication in progressive disease: patients' need to know and need to feel known. J Clin Oncol 2014 Nov 1;32(31):3474-3478. [doi: 10.1200/JCO.2014.56.0425] [Medline: 25267758]
8. Noordman J, Schulze L, Roodbeen R, et al. Instrumental and affective communication with patients with limited health literacy in the palliative phase of cancer or COPD. BMC Palliat Care 2020 Oct 7;19(1):152. [doi: 10.1186/s12904-020-00658-2] [Medline: 33028308]
9. Boykins AD. Core communication competencies in patient-centered care. ABNF J 2014;25(2):40-45. [Medline: 24855804]
10. Kwame A, Petrucka PM. A literature-based study of patient-centered care and communication in nurse-patient interactions: barriers, facilitators, and the way forward. BMC Nurs 2021 Sep 3;20(1):158. [doi: 10.1186/s12912-021-00684-2] [Medline: 34479560]
11. Cappella JN, Street RL. Delivering effective messages in the patient-clinician encounter. JAMA 2024 Mar 5;331(9):792-793. [doi: 10.1001/jama.2024.0371] [Medline: 38300603]
12. Zolnierek KBH, Dimatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. Med Care 2009 Aug;47(8):826-834. [doi: 10.1097/MLR.0b013e31819a5acc] [Medline: 19584762]
13. Cvetanovska N, Jessup RL, Wong Shee A, Rogers S, Beauchamp A. Patients' perspectives of factors influencing active participation in healthcare interactions: a qualitative study. Patient Educ Couns 2023 Sep;114:107808. [doi: 10.1016/j.pec.2023.107808] [Medline: 37263050]
14. Claramita M, Nugraheni MDF, van Dalen J, van der Vleuten C. Doctor-patient communication in Southeast Asia: a different culture? Adv Health Sci Educ Theory Pract 2013 Mar;18(1):15-31. [doi: 10.1007/s10459-012-9352-5] [Medline: 22314942]
15. Wang LY, Hu ZY, Chen HX, Zhou CF, Tang ML, Hu XY. Differences in regional distribution and inequality in health workforce allocation in hospitals and primary health centers in China: a longitudinal study. Int J Nurs Stud 2024 Sep;157:104816. [doi: 10.1016/j.ijnurstu.2024.104816]
16. Hu R, Liao Y, Du Z, Hao Y, Liang H, Shi L. Types of health care facilities and the quality of primary care: a study of characteristics and experiences of Chinese patients in Guangdong Province, China. BMC Health Serv Res 2016 Aug 2;16(a):335. [doi: 10.1186/s12913-016-1604-2] [Medline: 27484465]
17. Chen C, Liao M, Walther JB, Sundar SS. When an AI doctor gets personal: the effects of social and medical individuation in encounters with human and AI doctors. Commun Res 2024 Oct;51(7):747-781. [doi: 10.1177/00936502241263482]
18. Ruksakulpiwat S, Thorngthip S, Niyomyart A, et al. A systematic review of the application of artificial intelligence in nursing care: where are we, and what's next? J Multidiscip Healthc 2024;17:1603-1616. [doi: 10.2147/JMDH.S459946] [Medline: 38628616]
19. Braun V, Clarke V. Thematic Analysis: A Practical Guide: Sage Publications; 2021.
20. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
21. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
22. Chen Y, Zhang R, Lou Y, Li W, Yang H. Facilitators and barriers to the delivery of palliative care to patients with Parkinson's disease: a qualitative study of the perceptions and experiences of stakeholders using the socio-ecological model. BMC Health Serv Res 2023 Mar 6;23(1):215. [doi: 10.1186/s12913-023-09203-2] [Medline: 36879235]
23. Braun V, Clarke V. Toward good practice in thematic analysis: avoiding common problems and be(com)ing a knowing researcher. Int J Transgend Health 2023;24(1):1-6. [doi: 10.1080/26895269.2022.2129597] [Medline: 36713144]
24. Byrne D. A worked example of Braun and Clarke's approach to reflexive thematic analysis. Qual Quant 2022 Jun;56(3):1391-1412. [doi: 10.1007/s11135-021-01182-y]




XSL•FO
RenderX

25. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
26. Lincoln YS, Guba EG. Naturalistic Inquiry: SAGE; 1985.
27. Varpio L, Ajjawi R, Monrouxe LV, O'Brien BC, Rees CE. Shedding the cobra effect: problematising thematic emergence, triangulation, saturation and member checking. Med Educ 2017 Jan;51(1):40-50. [doi: 10.1111/medu.13124] [Medline: 27981658]
28. Street RL Jr. Communication in medical encounters: an ecological perspective. In: The Routledge Handbook of Health Communication: Routledge; 2003.
29. Herd P, Moynihan D. Health care administrative burdens: centering patient experiences. Health Serv Res 2021 Oct;56(5):751-754. [doi: 10.1111/1475-6773.13858] [Medline: 34515996]
30. Shanafelt TD. Physician Well-being 2.0: where are we and where are we going? Mayo Clin Proc 2021 Oct;96(10):2682-2693. [doi: 10.1016/j.mayocp.2021.06.005] [Medline: 34607637]
31. Mennella C, Maniscalco U, De Pietro G, Esposito M. Ethical and regulatory challenges of AI technologies in healthcare: a narrative review. Heliyon 2024 Feb 29;10(4):e26297. [doi: 10.1016/j.heliyon.2024.e26297] [Medline: 38384518]
32. Inui TS, Carter WB. Problems and prospects for health services research on provider-patient communication. Med Care 1985 May;23(5):521-538. [doi: 10.1097/00005650-198505000-00013] [Medline: 4010349]
33. Rao JK, Anderson LA, Inui TS, Frankel RM. Communication interventions make a difference in conversations between physicians and patients: a systematic review of the evidence. Med Care 2007 Apr;45(4):340-349. [doi: 10.1097/01.mlr.0000254516.04961.d5] [Medline: 17496718]
34. Verlinde E, De Laender N, De Maesschalck S, Deveugele M, Willems S. The social gradient in doctor-patient communication. Int J Equity Health 2012 Mar 12;11:12. [doi: 10.1186/1475-9276-11-12] [Medline: 22409902]
35. Olson DP, Windish DM. Communication discrepancies between physicians and hospitalized patients. Arch Intern Med 2010 Aug 9;170(15):1302-1307. [doi: 10.1001/archinternmed.2010.239] [Medline: 20696951]
36. Tu J, Kang G, Zhong J, Cheng Y. Outpatient communication patterns in a cancer hospital in China: a qualitative study of doctor-patient encounters. Health Expect 2019 Jun;22(3):594-603. [doi: 10.1111/hex.12890] [Medline: 30957939]
37. Street RL, Gordon H, Haidet P. Physicians' communication and perceptions of patients: is it how they look, how they talk, or is it just the doctor? Soc Sci Med 2007 Aug;65(3):586-598. [doi: 10.1016/j.socscimed.2007.03.036] [Medline: 17462801]
38. Cavalier JS, Goldstein BA, Ravitsky V, et al. Ethics in patient preferences for artificial intelligence-drafted responses to electronic messages. JAMA Netw Open 2025 Mar 3;8(3):e250449. [doi: 10.1001/jamanetworkopen.2025.0449] [Medline: 40067301]
39. Umar M, Ali V, Shamim L, et al. Transforming healthcare with large language models: current applications, challenges, and future directions—a literature review. J Intell Med 2026 Mar;3(1):8-25 [FREE Full text] [doi: 10.1002/jim4.70015]


## Abbreviations

**AI:** artificial intelligence
**COREQ:** Consolidated Criteria for Reporting Qualitative Research


*Edited by M Balcarras; submitted 17.Nov.2025; peer-reviewed by A Kidszun, MA Nas; revised version received 14.Apr.2026; accepted 27.Apr.2026; published 11.Jun.2026.*

*Please cite as:*
*Shi Z, Jiang Q, Wang X, Yan H, Xia Y, Shi L, Wang D*
*Barriers and Facilitators to Physician-Patient Communication in Chinese Tertiary Hospitals From the Perspectives of Hospital-Based Physicians and Patient Relations Coordinators: Qualitative Study*
*Interact J Med Res 2026;15:e87947*
*URL: https://www.i-jmr.org/2026/1/e87947*
*doi:10.2196/87947*


© Zengping Shi, Qinqin Jiang, Xincheng Wang, Hongli Yan, Yi Xia, Lushaobo Shi, Dong Wang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 11.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Codeveloping a Clinical Pathway for Recurrent Clostridioides difficile Infection With Primary Care Providers: User-Centered Design Study

Ritika Kamlesh Patel[1], MPH; Rose Franz[1], BScN, RN; Jordan Tate[2], BDes; Jennifer Coppens[3], MD; Matthew Do Rego[3], MD; Christopher Smith[3], MD; Lauretta Ovadje[4], PhD; Karen Wong[3], MD; Lynora Saxinger[3], MD; Dina Kao[1], MD

[1]Division of Gastroenterology, Department of Medicine and Dentistry, University of Alberta, 2-24 Zeidler Ledcor Center, 8540-112 St NW, Edmonton, AB, Canada

[2]Physician Learning Program, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada

[3]Department of Family Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada

[4]Health Systems Knowledge and Evaluation, Acute Care Alberta, Edmonton, AB, Canada

**Corresponding Author:**
Dina Kao, MD
Division of Gastroenterology, Department of Medicine and Dentistry, University of Alberta, 2-24 Zeidler Ledcor Center, 8540-112 St NW, Edmonton, AB, Canada

## Abstract

**Background:** *Clostridioides difficile* infection (CDI) is the most common cause of health care–associated infectious diarrhea; however, variability in laboratory testing results in uncertainty in diagnosis and clinical decision-making. Recurrent CDI (rCDI), defined as the return of infection following initial treatment, continues to present significant clinical challenges due to limited therapeutic options. Although health care providers are familiar with vancomycin as first-line therapy for CDI, many are less familiar with fecal microbiota transplantation (FMT), despite its recommendation in clinical guidelines for rCDI management. Consequently, primary care providers (PCPs) may experience uncertainty regarding diagnostic test interpretation, treatment selection, and referral for FMT, highlighting the need for practical clinical tools to support evidence-based management.

**Objective:** This study will support the standardization of care and improve patient treatment experiences in Alberta by identifying barriers and facilitators encountered by PCPs in managing CDI and rCDI, using these findings to codevelop a provincial rCDI clinical pathway.

**Methods:** This qualitative study investigated the codevelopment process of a provincial rCDI clinical pathway between the study team and intended end users. A draft pathway was developed by the study team comprising gastroenterology and infectious disease specialists, followed by an iterative end user–centered design approach. PCP participants completed 2 rounds of 60 - to 90-minute virtual, one-on-one, semistructured interviews incorporating scenario-based usability testing with standardized clinical cases. Using a think-aloud approach, participants evaluated the pathway's clarity, diagnostic decision-making, treatment recommendations, referral processes, and integration into clinical workflows. Feedback was analyzed using thematic analysis, informed iterative revisions were made to the pathway and accompanying user guide, and they were reevaluated in a second round of interviews.

**Results:** Out of 30 invited PCPs, 3 (10%) participated in this study. During the first round of interviews, participants identified challenges interpreting PCR-positive vs toxin-positive test results, navigating the pathway layout, locating embedded hyperlinks, and understanding FMT referral processes. Iterative, feedback-driven revisions addressed these barriers by improving diagnostic guidance, simplifying navigation, increasing hyperlink visibility, and clarifying referral timing with a user guide. During the second round of interviews, all participants reported that modifications improved CDI test result interpretation, streamlined navigation, enhanced workflow integration, and increased confidence in applying the pathway to clinical scenarios. The final version of this pathway was well received for its clarity, readability, and clinical utility, and was viewed as a valuable educational resource for supporting evidence-based management of CDI.

**Conclusions:** This study demonstrated the feasibility and value of codeveloping a provincial clinical pathway with physician partners for managing CDI and rCDI. The resulting tool was clear, practical, and aligned with primary care workflows. This iterative, user-centered design was useful for developing clinically relevant decision support tools. Future research should evaluate the pathway's implementation and impact on physician practice, patient care, and clinical outcomes using quantitative and qualitative methods.



XSL•FO
RenderX

*(Interact J Med Res 2026;15:e93032)* doi:10.2196/93032

**KEYWORDS**

Clostridioides difficile infection; clinical pathway; codevelopment; primary care providers; fecal microbiota transplant

## Introduction

*Clostridioides difficile* infection (CDI) is a significant health care priority, representing the most common causes of health care–associated infectious diarrhea, commonly triggered by antibiotic use [1]. Gaps in knowledge around CDI diagnosis and recurrent CDI (rCDI) management have contributed to suboptimal and delayed treatments [2,3]. Specifically, CDI is diagnosed in patients who have diarrhea and a positive stool test for toxigenic *Clostridioides difficile*. Diagnostic confusion arises because there are two main categories of laboratory assays: (1) an enzyme immunoassay for toxins A and/or B with high specificity but low sensitivity (33% - 65%) [4], and (2) a polymerase chain reaction (PCR)–based assay with higher sensitivity (approximately 90%) but much lower specificity, as it cannot differentiate between active infection and asymptomatic colonization [5]. While most practitioners are comfortable with vancomycin as the first-line therapy for CDI, most are not familiar with guideline-recommended fecal microbiota transplantation (FMT) for patients with at least 2 recurrences [6,7]. Not only is FMT an efficacious treatment option, it is also the most cost-effective strategy for treating rCDI [8]. In a retrospective study in Alberta, timely FMT, defined as administration after the second recurrence, was associated with mean per-patient cost savings of approximately CAD $29,842 (CAD $1=US $0.71 as of September 23, 2026) compared with delayed FMT [9]. Additional challenges exist as many primary care providers (PCPs) are unsure when to refer [10] or how to access FMT.

Clinical pathways are evidence-based tools designed to translate clinical practice guideline recommendations into routine practice while standardizing and coordinating patient care within a multidisciplinary health care setting [11]. A study evaluating the implementation of a clinical pathway for liver surgery found that its use significantly reduced hospital stays and postoperative costs without increasing postoperative morbidity [12]. However, traditional top-down dissemination of clinical practice guidelines often fails to address potential barriers to their development, dissemination, and implementation [13,14]. As such, it is crucial to involve end users in codeveloping this rCDI pathway to optimize guideline implementation and timely care for rCDI patients.

## Methods

### Study Overview

This study describes the process of cocreating a provincial clinical pathway to support physicians in the management of patients with CDI (including rCDI) in Alberta, Canada, while highlighting key aspects of CDI care, including diagnosis, treatment, and referral processes.

### Ethical Considerations

Ethics approval was obtained from the University of Alberta Research Ethics Board (Pro00130489). All participants signed a written consent form and agreed to be video recorded during the interviews.

### Physician Recruitment

A purposive sampling strategy was applied to recruit participants. PCPs were identified with the support of the Alberta Health Services (AHS) Primary Health Care Integration Network, interested primary care networks, the Alberta Medical Association, and the academic Department of Family Medicine at the University of Alberta. From this pool, PCPs with experience referring rCDI patients were contacted via email to assess their interest in participating. The FMT nurse navigator distributed consent forms and coordinated interviews.

### Study Design and Codevelopment Process

We used a qualitative methodology to gather empirical data on the optimal format and clinical content of the proposed pathway. The draft provincial rCDI clinical pathway was first developed by the study team consisting of specialists in gastroenterology and infectious diseases in collaboration with the University of Alberta's Physician Learning Program. To facilitate use in clinical practice, the pathway incorporated direct links to existing provincial and Canadian clinical resources, including the Alberta Referral Directory, Bugs & Drugs [15], and other relevant guidance documents, allowing clinicians to access additional information as needed. This draft served as a low-fidelity prototype intended for iterative testing and refinement according to feedback from the end users. A user-centered design approach was also taken, as scenario-based usability testing and iterative prototyping can reveal cognitive and workflow mismatches that are not readily apparent through guideline development alone. Codevelopment activities consisted of 2 rounds of scenario-based usability testing conducted through one-on-one virtual 60 - to 90-minute interview sessions with PCPs. The sessions followed a semistructured format, framed by an interview guide with case scenarios in addition to open-ended questions. A separate interview guide was designed for each respective interview.

Prior to the first interview, PCPs were provided with the initial draft of the provincial rCDI pathway. During the interview, participants were asked to apply the draft pathway to 2 standardized patient case scenarios representing common but clinically challenging rCDI presentations. Using a think-aloud approach, PCPs described how they interpreted pathway steps, made treatment decisions, and determined referral appropriateness while navigating the tool.

The interviewer probed for clarity, decision confidence, points of hesitation, perceived ambiguities, and workflow misalignments. Particular attention was paid to how physicians interpreted diagnostic test results (PCR vs toxin), distinguished



XSL•FO
RenderX

recurrent cases from new infections, and identified when to reassess for alternative causes of diarrhea.

Following the first interview, usability findings were synthesized into structured summary tables and task success matrices capturing decision points, areas of confusion, drop-off locations within the pathway, and suggested revisions. These artifacts were used to guide iterative pathway redesign, including modifications to terminology, visual hierarchy, navigation cues, hyperlink visibility, and referral timing.

The revised pathway and an accompanying user guide were then retested with the same end user PCPs in a second round of interviews to assess whether prior usability issues had been resolved and to validate improvements in clarity, workflow integration, and clinical feasibility.

### Data Collection and Analysis

The AHS Health System Knowledge and Evaluation team, acting as additional independent qualitative analysts, collaborated with the research team to design interview guides including sample case scenarios, facilitate interviews, and conduct data analysis. Verbatim transcripts generated from Microsoft Teams recordings were analyzed using a thematic analysis framework. To ensure methodological rigor, coding was conducted independently by a member of the study team and a member of the external AHS Health System Knowledge and Evaluation team. Discrepancies in code applications were resolved through consensus and consultation with the broader study team rather than using formal quantitative interrater reliability measures.

Theme derivation followed an inductive approach, progressing from initial open coding to group codes into broader, descriptive categories representing specific workflow barriers or design challenges. These categories were then structured into overarching thematic maps (Figures S1 and S2 in Multimedia Appendix 1) to visually map the relationships between diagnostic interpretation, pathway navigation, clinical workflow, and implementation feasibility.

In addition to thematic qualitative analysis, usability data were analyzed using task-based evaluation methods commonly used in human-centered design. Task success and failure were documented alongside qualitative observations and participant quotations. This dual analytic approach enabled triangulation between what physicians reported, how they navigated the pathway, and where design-related barriers emerged. Findings were iteratively mapped back to pathway revisions to ensure that identified usability issues directly informed design changes rather than remaining descriptive only.

## Results

### Participants

Of the 30 PCPs invited, 3 provided consent and subsequently participated in one-on-one interviews. Participant gender and practice history are presented in Table 1.

**Table .** Participant characteristics (N=3).

| Characteristic | Participants, n (%) |
|---|---|
| Male | 2 (66.7) |
| Health care specialty and practice setting | |
| Primary care | 3 (100) |
| Plus emergency medicine | 1 (33.3) |
| Plus geriatrics | 1 (33.3) |
| Urban | 2 (66.7) |
| CDI[a] patient volume (per year) | |
| <10 patients | 2 (66.7) |
| >10 patients | 1 (33.3) |

[a]CDI: *Clostridioides difficile* infection.

### Interview Sessions

#### *PCP Interviews Round 1*

Three PCPs participated in 2 rounds of semistructured, one-on-one interviews. All 3 participants worked in primary care settings with annual patient volumes exceeding 500.

The first set of interviews were conducted between May 29, 2024, and June 6, 2024. The initial draft of the pathway incorporated direct access links to several resources: (1) Bugs & Drugs, an evidence-based antimicrobial reference guide for health care professionals in Alberta [15]; (2) the AHS diarrhea pathway, a clinical guide for managing diarrhea in primary care, including a step-by-step approach with investigations and management; and (3) patient education resources. A warning sign–flagged hyperlink was incorporated to distinguish fulminant CDI from severe nonfulminant CDI and to direct clinicians to the inpatient CDI management pathway for cases requiring urgent care. The pathway begins by outlining clinical criteria that define “suspected new or recurrent CDI,” followed by a recommendation to obtain a *C difficile* diagnostic test in the appropriate context. Subsequent management is guided by the test results, which may be (1) toxin positive, (2) PCR positive with additional risk assessment, or (3) both toxin and PCR negative. Based on these findings, the pathway specifies what treatment is required and outlines the recommended

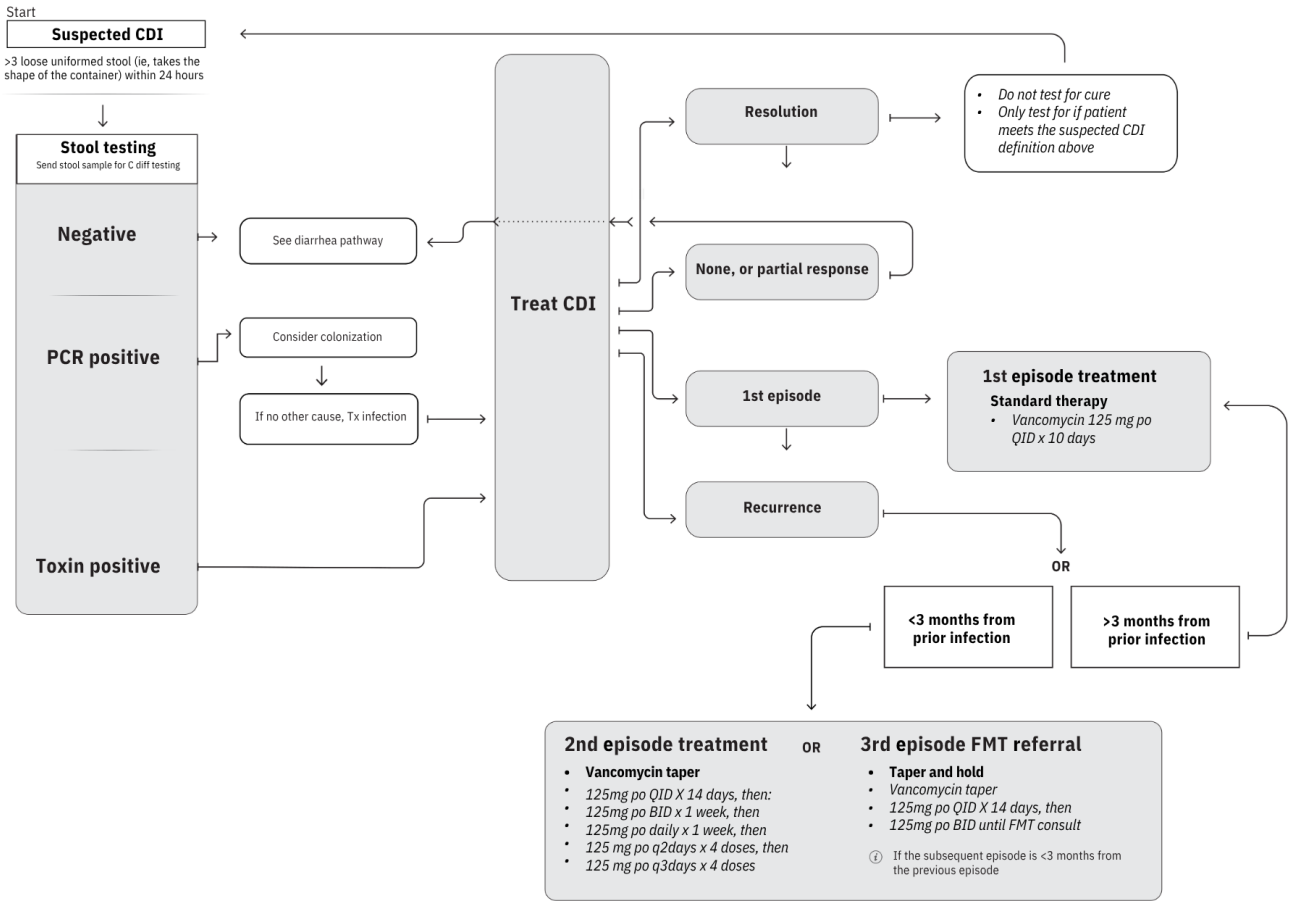


XSL•FO
RenderX

clinical actions for CDI recurrence, including management for first, second, and third episodes. The pathway explicitly indicates FMT referral at the third episode (or second recurrence). Further recommendations are based on treatment response.

The initial and final versions of the pathway can be found in Figures 1 and 2, respectively. Visually, each major phase of decision-making is separated by dotted lines to delineate distinct stages of progression. Individual outcomes are presented in separate boxes to emphasize branching pathways. Embedded hyperlinks within the outcome boxes direct users to supplementary informational resources when required.

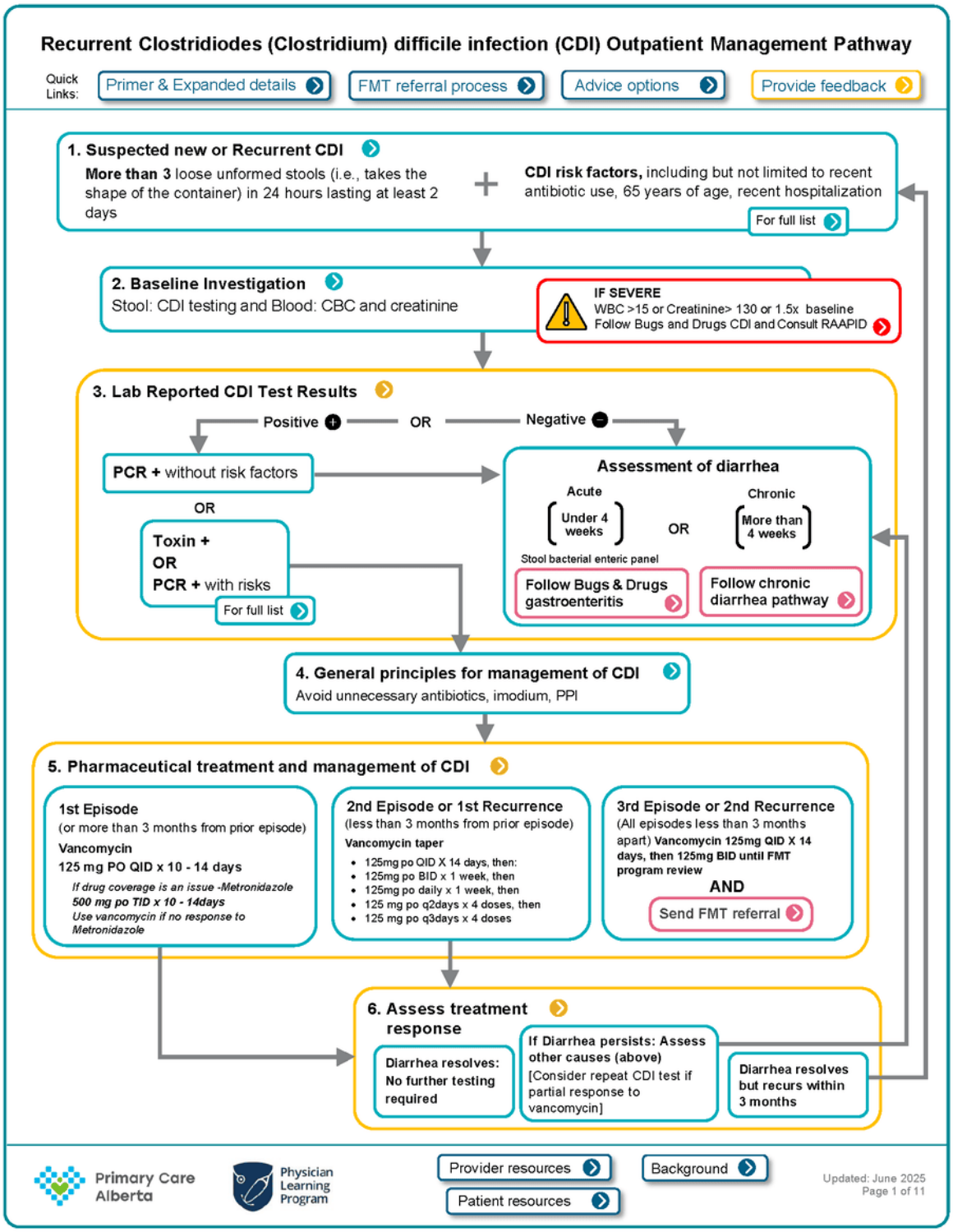
**Figure 1.** Initial draft of the recurrent *Clostridiodes difficile* infection (rCDI) clinical pathway prior to finalization. BID: twice a day (*bis in die*); C diff: *Clostridiodes difficile*; CDI: *Clostridiodes difficile* infection; FMT: fecal microbiota transplantation; PCR: polymerase chain reaction; po: by mouth (*per os*); q2days: every 2 days (*quaque* 2 days); q3days: every 3 days (*quaque* 3 days); QID: 4 times a day (*quater in die*); Tx: treatment.






XSL•FO
RenderX

**Figure 2.** Final version of the recurrent *Clostridiodes difficile* infection (rCDI) clinical pathway. BID: twice a day (*bis in die*); CBC: complete blood count; FMT: fecal microbiota transplantation; PCR: polymerase chain reaction; po: by mouth (*per os*); PPI: proton pump inhibitor; q2days: every 2 days (*quaque* 2 days); q3days: every 3 days (*quaque* 3 days); QID: 4 times a day (*quater in die*); RAAPID: referral, access, advice, placement, information, destination; WBC: white blood cell.




The first interview focused on eliciting participant feedback on the clarity and usability of the pathway. Participants were asked to apply the pathway to 2 patient scenarios and describe how they would manage each case, including the need for FMT referral and the appropriate timing and process for referral.

Task-based usability outcomes from the scenario-based interviews are summarized in Table 2. Participants completed a series of predefined clinical decision-making tasks, including distinguishing between new and recurrent CDI, interpreting diagnostic test results, selecting the appropriate management pathway, identifying the appropriate timing for FMT referral, and navigating the pathway to relevant supporting resources.



XSL•FO
RenderX

Areas of task difficulty identified during the first round of usability testing informed iterative revisions to the pathway prior to the second round of interviews.

**Table .** Participant performance on predefined usability tasks during pathway evaluation (N=3).

| Usability task | Round 1 | Round 2 |
|---|---|---|
| Accurate classification of recurrence vs new infection | 2/3 success; 1/3 partial success[a] | 3/3 success |
| Interpretation of PCR[b]-positive vs toxin-positive results | 0/3 success; 3/3 partial success[a] | 3/3 success |
| Appropriate treatment selection | 3/3 success | 3/3 success |
| Correct identification of FMT[c] referral thresholds | 2/3 success; 1/3 partial success[a] | 3/3 success |
| Confidence[d] in next-step decision-making | 2/3 success; 1/3 partial success[a] | 3/3 success |

[a]Partial success means the participant ultimately reached an appropriate decision but required clarification, showed significant hesitation, or completed only 1 of the 2 scenarios correctly.

[b]PCR: polymerase chain reaction.

[c]FMT: fecal microbiota transplantation.

[d]Confidence was not measured using a numerical scale during the interviews. This row is based on retrospective coding of participant hesitation, requests for clarification, and ability to independently identify the appropriate next step.

Complementary qualitative findings are presented in Figure S1 in Multimedia Appendix 1, which illustrates the thematic map generated from the first round of interviews. The thematic analysis identified barriers related to pathway design, structural organization, diagnostic test interpretation, navigation, and referral processes. Together, the usability and qualitative findings informed iterative refinements to improve the pathway's clarity, usability, and alignment with practitioner workflows.

> *I guess one of my questions would be about the testing. I think there needs to be something on this path...that talks about the difference between toxin testing and PCR testing, as I suspect I am not alone as a family physician in wondering when each of those happens.* [Participant 3]

> *That is a little bit ambiguous in terms of the PCR testing, that I would...clean that up, but just so that it makes it a little bit more clear cut that that's what you mean...I understand now that it's explained, but I wouldn't have thought that straight away.* [Participant 2]

Participants reported difficulty navigating the draft pathway, as the dotted lines gave the impression that sections were separate and did not allow users to circle back through the algorithm. Consequently, PCPs were uncertain about when to assess patients for alternative causes of diarrhea (Figure 1).

> *I'm wondering if...there's a way to link back from...this algorithm...it kind of makes it look like once you're in the "treat C diff" you don't go back, although I guess down at the bottom under treatment outcome you go down to "Assess other causes".* [Participant 1]

PCPs recommended clearer visual cues for hyperlinks, such as the use of consistent icons to indicate their presence. While the draft pathway clearly indicated when patients should be referred for FMT, 2 PCPs noted concerns about the referral process itself. Additionally, the placement of the FMT referral button did not clearly convey if the referral should occur concurrently with vancomycin treatment.

> *When I kind of first looked at the pathway it just kind of seemed a little bit out of place. Like I didn't know what, like where that kind of related, if you don't click the links.* [Participant 1]

> *Just in case there is anyone that perceives that as an ambiguity...If you send the FMT referral, if you put that as the first thing, and then you do that concurrently.* [Participant 2]

All participating PCPs reported that the pathway enabled them to distinguish between a new CDI infection and a recurrent infection. They also indicated that the pathway clearly outlined the treatment approach for both scenarios. Participants recommended including an appendix or user guide containing definitions and additional information to avoid having to consult an external source.

> *So on a lot of the specialist link pathways and provincial pathways and stuff they have the page like this, and then there's a link within here that just leads to almost an appendix that it details out risk factors or details out like higher risk features in more detail kind of things so that I don't have to go to a second source.* [Participant 1]

PCPs were asked whether their treatment approach differed when a patient's results were PCR positive vs toxin positive, and all indicated that it did. PCPs considered a PCR-positive result in patients without risk factors indicative of colonization and did not initiate CDI treatment. In cases where patients showed no response or a partial response to CDI therapy, PCPs reported extending treatment duration, conducting additional testing for alternative causes, and referring the patient as appropriate.

> *I would do more of a work up...broaden my differential and look at things again. If they were toxin-positive and didn't respond...I know the toxin*



XSL•FO
RenderX

> *takes a while to go away, even with treatment. So it would kind of depend how long, I might consider retesting for C diff, and...if the toxin was negative, [it] would be more helpful than if it's positive. But ultimately if I'm just really not sure what to do, I would usually do a connect MD phone consult or something like that.* [Participant 1]

When asked about the potential use of patient education materials linked to the pathway, all participating PCPs indicated they would use such resources. Additional feedback highlighted variability in clinical workflows, including differences between practices using provincially based electronic medical records or paper charts, which may affect how this clinical pathway is accessed. Some PCPs noted that certain pathway content was not fully aligned with information in Bugs & Drugs, which highlighted the need for consistency with existing reference materials.

> *That could be a very big hindrance in following this pathway [the misalignment with Bugs & Drugs]...I've dealt with some really recurrent C diff so I don't put people on Metronidazole. I put them on Vanco as my rule of thumb...whenever I have students and I teach them this because I'm also a precept at the university. I try and explain this and they said "no, no, but Bugs & Drugs says..." and I said "yeah, that's fine." But that's not what we should do...* [Participant 2]

Based on participant feedback, we incorporated additional resources, including links to alternative pathways for varying clinical scenarios and detailed information on the FMT referral process. The revised version was then presented to the same physicians during a second interview to further refine the pathway.

### *PCP Interviews Round 2*

The second round of interviews was conducted between February 27, 2025, and March 11, 2025. Prior to these interviews, PCPs were provided with the revised provincial rCDI clinical pathway along with the user guide, which had been developed based on feedback from the first round of interviews.

Figure S2 in Multimedia Appendix 1 presents the thematic map derived from analysis of the second interview, focused on evaluating the revised pathway and assessing whether ambiguities were resolved. Overarching themes reflected improved diagnostic clarity, navigation, referral processes, and considerations related to workflow integration and implementation.

Task-based usability testing demonstrated improvements across all predefined tasks following pathway revision (Table 2), supporting the qualitative findings that participants experienced greater clarity in diagnostic interpretation, referral processes, and pathway navigation.

> *It was much more readable and more consistent with the other pathways.* [Participant 1]

The participating PCPs indicated that the revised pathway had improved clarity in diagnostic test interpretation, differentiation between new and recurrent CDI, referral timing, and overall ease of navigation.

> *It [the pathway] does [clarify test results]. Just the addition of the two categories on the third step is useful.* [Participant 3]
>
> *I think it makes it more clear about what's a recurrence versus what's a new infection...Yeah, I thought that section was definitely clarified. For sure, and easier to read.* [Participant 1]

One participant suggested removing the "If severe" box because the guidance provided in Bugs & Drugs was consistent with the information already in the pathway.

> *Because presumably, if your white blood cells are high and your creatinine's high, you also need the positive CDI test...if it's C. diff, then you go down to five. And it doesn't change what you do for that first episode, apart from the clinical assessment of maybe they need hydration or admission or something like that...but the instructions to follow Bugs & Drugs CDI, you're kind of doing that anyway. Going through the pathway, so you don't need to shunt someone else to a different source when it's exactly the same as what's later on in the pathway.* [Participant 1]

The user guide (Multimedia Appendix 2) was perceived as a comprehensive supplement to the pathway and was particularly helpful in diagnostic test interpretation, including the distinction between PCR-positive and toxin-positive results. However, participants indicated that the user guide added little additional value for CDI management or differentiating new vs recurrent infections, as the pathway itself already provided sufficient clarity.

> *That's [the User Guide] exactly what provides the clarity for sure.* [Participant 2]
>
> *Truthfully, it's pretty well laid out in the pathway. So really, in terms of the treatment itself, I don't know that the User Guide contributes a lot to the treatment component of it specifically...given that the pathway's quite clear on what the treatment would and should be.* [Participant 3]

This sample of PCPs viewed the pathway as a valuable educational resource and recommended several refinements to the accompanying user guide, including guidance for managing patients with a vancomycin allergy and providing contact information for physicians requiring further support. A recurring theme across interviews was how and when the pathway would be used. One suggestion was to attach or link the pathway to a positive CDI laboratory result, ensuring that physicians have the provincial rCDI clinical pathway and user guide available at the point of care.

> *That might be something to put on the PCR results of C diff testing at the bottom. Say, "please see Provincial Pathway for CDI, for recurrent CDI management." Because a lot of times someone comes in with diarrhea, you know, I think a lot of people are like, OK, C diff, I'm going to check for it and then they get the results back and they're like, "what do I*



XSL•FO
RenderX

> *do with this now?" And if you have the link directly there in the results...they're much more likely to look at it.* [Participant 1]
>
> *It's a good teaching tool in that sense, to physicians as well, because that will be implicit if someone follows the pathway and follows that course...it helps teach for the future.* [Participant 3]

Given the small sample size, thematic saturation may have been limited. However, recurring themes were consistently identified across interviews, supporting the credibility of the findings within this exploratory sample.

## Discussion

### Principal Result

Physician partners in our study identified important gaps in the current CDI clinical workflow, including inconsistencies in referral patterns and challenges in accessing consistent information. Through an iterative co-creation process with PCPs, we developed a physician-centered pathway to guide diagnosis, management, and referral for rCDI. The resulting pathway was perceived by participants as clinically relevant, practical, and easy to use, highlighting the feasibility of using a co-creation framework to design practical, end user–aligned clinical pathways. However, our participating physicians were all highly engaged, academically affiliated, and had prior experience referring rCDI patients, and results may differ if we had included participants in rural practices or those without prior experience with rCDI patients. Nevertheless, as a regional, qualitative development study, our results demonstrate the value of the co-design process. Future studies should evaluate pathway uptake and impact on clinical practice, referral patterns, health care utilization, and patient outcomes.

### Current Landscape

Prior to this project, no provincial clinical pathway existed for CDI management in Alberta. While CDI practice guidelines have been published by professional societies, access to FMT in Canada remains limited, and most practitioners do not know when or how to access this treatment [16]. Clinical pathways may address these gaps by translating evidence into practice and improving access to emerging treatments. Integration of pathway content into electronic health records has been proposed as a low-effort strategy to enhance awareness [17], which informed the development of our clinical pathway.

### Co-Creation Process

Co-creation is gaining recognition as a valuable way to bring together stakeholders to promote collaboration, support knowledge dissemination, and exchange perspectives [18]. In practice, coordinating multiple health care professionals during pathway development can be challenging. The complexity of the clinical problem, intervention, and implementation, including the need to align pathway goals across members of the health care team and ensure stakeholder engagement, may either facilitate or hinder both pathway development and implementation [19]. Also, a lack of leadership and limited intrinsic motivation can be barriers in a clinical setting [19]. Therefore, involving a specialist in a relevant health care field is essential for providing guidance and structure in the initial development of care pathways [20]. Across specialties, pathways have demonstrated effectiveness in standardizing patient care, improving timeliness, and reducing costs while not negatively affecting overall patient health outcomes [12,21]. For effective knowledge mobilization, particularly between specialists and PCPs, codevelopment provides an important mechanism for generating tools that are relevant and acceptable to stakeholders, more usable, and more likely to be successfully implemented within health care settings [22,23].

Embedding human-centered design methods allowed us to translate physician feedback into iterative refinements that improved our pathway's clarity and usability. Although the initial draft was developed by the multidisciplinary research team using existing evidence and specialist expertise, repeated engagement with PCPs allowed their perspectives to inform iterative pathway refinement. Future work could involve PCPs earlier in pathway conceptualization, enabling them to shape pathway content from the outset while further strengthening stakeholder ownership and supporting implementation.

### Clinical Pathway Efficacy

The existing literature examining the co-creation of clinical pathways with PCPs is limited. However, studies evaluating clinical pathways more broadly demonstrate that their implementation can yield numerous positive outcomes across the journey of care, from diagnosis to treatment. One study assessing the impact of a clinical pathway reported a marked reduction in health care–associated CDI rates following the introduction of a mandatory pathway [24]. The investigators attributed this reduction to decreased rates of misdiagnosis, reducing the use of resources and health care costs and promoting the delivery of appropriate patient care.

Evidence from other studies further emphasizes the utility of clinical pathways in other areas. A study conducted in the United States found that implementing clinical pathways for nutritional management and lower airway inflammation in patients with cystic fibrosis resulted in more standardized and uniform care [25]. Additionally, a systematic review from Europe reported that care pathways enhanced health care staff knowledge, improved interprofessional documentation, and strengthened both communication and team relationships [26]. Across these studies, a recurring theme was the necessity of a multidisciplinary approach in developing and implementing care pathways, which can effectively support collaboration within interprofessional teams and enhance the quality of patient care [25,26].

### Limitations

A key limitation in this study is the low number of participants. Potential participants may have been discouraged by the required time commitment of up to 4 hours, which can be a significant challenge for a busy clinician. Selection bias may also have been introduced in this exploratory codevelopment study; as participants were recruited through professional and academic networks, all 3 PCPs were academically affiliated and had prior experience treating and referring rCDI patients. Self-selection bias was also a concern, as those who agreed to participate may



XSL•FO
RenderX

have held stronger opinions regarding clinical pathways or a greater interest in quality improvement initiatives. Although this small sample size was appropriate for iterative pathway development, the findings may not fully reflect the perspectives of physicians practicing in other settings, those who have not managed rCDI patients, or other potential end users such as nurses, pharmacists, or patients. A greater diversity of participants may have provided valuable perspectives for further refining this pathway and its implementation. A higher number of participants would also permit specific feedback to be quantified to substantiate our qualitative analyses. As such, the findings should be viewed as exploratory and may not be transferable to community-based, rural-based, or less academically engaged practitioners and allied health care team members.

Supporting PCP engagement in research is essential to developing systemic interventions that mitigate clinical burdens and foster robust, interdisciplinary support networks across medical specialties [27,28]. Enhancing future PCP research participation requires a multipronged approach, which may include equitable financial compensation, reduced interference with daily practice routines, and enhanced institutional backing [29,30]. Future pathway refinement should also incorporate perspectives from additional stakeholders, including nurses, pharmacists, and a more diverse group of physicians, to strengthen the pathway implementation and transferability.

Methodological limitations should also be noted. The interview guides were not pilot tested, potentially limiting our ability to assess and refine the validity of their content [31]. Interviews are a valuable method for gathering information; however, the quality of data obtained through interviews can vary considerably depending on the researcher's qualifications, skills, and familiarity with the research topic.

## Conclusions

The co-creation of clinical pathways with physician partners represents a feasible approach for developing clinically relevant decision support tools for the management of rCDI. Future research should explore the acceptability and effectiveness of alternative formats for the pathway to ensure these resources remain accessible and relevant across diverse primary care settings. As the rCDI clinical pathway is introduced into the provincial referral directory, prospective evaluation will be critical to assess its impact on clinical decision-making, referral practices, implementation, and patient outcomes, while informing future refinement of the pathway.

## Acknowledgments

No generative AI or AI-assisted technologies were used in the preparation of this manuscript.

## Funding

Funding for this study was provided by the Digestive Health Strategic Clinical Network of Alberta Health Services and supported by the Physician Learning Program. The Physician Learning Program is funded by the Government of Alberta. The views expressed herein do not necessarily represent the official policy of the Government of Alberta.

## Data Availability

Data is available upon request to the corresponding author (DK).

## Authors' Contributions

Conceptualization: RKP, RF, JT, KW, LS, DK
Formal analysis: RKP, RF, JT, KW, LS, DK
Funding acquisition: KW, DK
Investigation: RKP, RF, JT, JC, MDR, CS, KW, LS, DK
Methodology: RKP, RF, KW, DK
Visualization: RKP, RF, JT, JC, MDR, CS, LO, KW, DK
Writing – original draft: RKP, RF, KW, DK
Writing – review & editing: RKP, RF, JT, JC, MDR, CS, LO, KW, LS, DK

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures 1 and 2 of physician interview thematic maps.
[ZIP File, 151 KB - i-jmr_v15i1e93032_app1.zip ]

Multimedia Appendix 2
Pathway user guide.
[ZIP File, 236 KB - i-jmr_v15i1e93032_app2.zip ]



XSL•FO
RenderX

## References


1. Feuerstadt P, Theriault N, Tillotson G. The burden of CDI in the United States: a multifactorial challenge. BMC Infect Dis 2023 Mar 7;23(1):132. [doi: 10.1186/s12879-023-08096-0] [Medline: 36882700]
2. Markantonis JE, Fallon JT, Madan R, Alam MZ. Clostridioides difficile infection: diagnosis and treatment challenges. Pathogens 2024 Jan 27;13(2):118. [doi: 10.3390/pathogens13020118] [Medline: 38392856]
3. Lev V, Anbarchian T, Yao H, Bhat A, Britt P, Shieh L. Health care–associated Clostridioides difficile infection: learning the perspectives of health care workers to build successful strategies. Am J Infect Control 2024 Mar;52(3):284-292. [doi: 10.1016/j.ajic.2023.08.008] [Medline: 37579972]
4. Humphries RM, Uslan DZ, Rubin Z. Performance of Clostridium difficile toxin enzyme immunoassay and nucleic acid amplification tests stratified by patient disease severity. J Clin Microbiol 2013 Mar;51(3):869-873. [doi: 10.1128/JCM.02970-12] [Medline: 23269736]
5. Deshpande A, Pasupuleti V, Rolston DDK, et al. Diagnostic accuracy of real-time polymerase chain reaction in detection of Clostridium difficile in the stool samples of patients with suspected Clostridium difficile infection: a meta-analysis. Clin Infect Dis 2011 Oct;53(7):e81-e90. [doi: 10.1093/cid/cir505] [Medline: 21890762]
6. Peery AF, Kelly CR, Kao D, et al. AGA clinical practice guideline on fecal microbiota-based therapies for select gastrointestinal diseases. Gastroenterology 2024 Mar;166(3):409-434. [doi: 10.1053/j.gastro.2024.01.008] [Medline: 38395525]
7. Kelly CR, Fischer M, Allegretti JR, et al. ACG clinical guidelines: prevention, diagnosis, and treatment of Clostridioides difficile infections. Am J Gastroenterol 2021 Jun 1;116(6):1124-1147. [doi: 10.14309/ajg.0000000000001278] [Medline: 34003176]
8. Konijeti GG, Sauk J, Shrime MG, Gupta M, Ananthakrishnan AN. Cost-effectiveness of competing strategies for management of recurrent Clostridium difficile infection: a decision analysis. Clin Infect Dis 2014 Jun;58(11):1507-1514. [doi: 10.1093/cid/ciu128] [Medline: 24692533]
9. Waye A, Atkins K, Kao D. Cost averted with timely fecal microbiota transplantation in the management of recurrent Clostridium difficile infection in Alberta, Canada. J Clin Gastroenterol 2016 Oct;50(9):747-753. [doi: 10.1097/MCG.0000000000000494] [Medline: 26890327]
10. Jackson M, Olefson S, Machan JT, Kelly CR. A high rate of alternative diagnoses in patients referred for presumed Clostridium difficile infection. J Clin Gastroenterol 2016 Oct;50(9):742-746. [doi: 10.1097/MCG.0000000000000447] [Medline: 26565971]
11. Mohan S, Nandi D, Stephens P, M Farrej M, Vogel RL, Bonafide CP. Implementation of a clinical pathway for chest pain in a pediatric emergency department. Pediatr Emerg Care 2018 Nov;34(11):778-782. [doi: 10.1097/PEC.0000000000000861] [Medline: 27649041]
12. Ovaere S, Boscart I, Parmentier I, et al. The effectiveness of a clinical pathway in liver surgery: a case-control study. J Gastrointest Surg 2018 Apr;22(4):684-694. [doi: 10.1007/s11605-017-3653-1] [Medline: 29274000]
13. Wang T, Tan JYB, Liu XL, Zhao I. Barriers and enablers to implementing clinical practice guidelines in primary care: an overview of systematic reviews. BMJ Open 2023 Jan 6;13(1):e062158. [doi: 10.1136/bmjopen-2022-062158] [Medline: 36609329]
14. Guerra-Farfan E, Garcia-Sanchez Y, Jornet-Gibert M, Nuñez JH, Balaguer-Castro M, Madden K. Clinical practice guidelines: the good, the bad, and the ugly. Injury 2023 May;54 Suppl 3:S26-S29. [doi: 10.1016/j.injury.2022.01.047] [Medline: 35135686]
15. Blondel-Hill E, Fryters S. Bugs & Drugs. URL: https://www.bugsanddrugs.org/ [accessed 2026-09-25]
16. Hocking L, Ianiro G, Leong RW, et al. Faecal microbiota transplantation for recurrent C. difficile infections: challenges and improvement opportunities for clinical practice and healthcare systems. Aliment Pharmacol Ther 2023 Mar;57(5):549-564. [doi: 10.1111/apt.17309] [Medline: 36495561]
17. Flores EJ, Jue JJ, Giradi G, Schoelles K, Mull NK, Umscheid CA. AHRQ EPC series on improving translation of evidence: use of a clinical pathway for C. difficile treatment to facilitate the translation of research findings into practice. Jt Comm J Qual Patient Saf 2019 Dec;45(12):822-828. [doi: 10.1016/j.jcjq.2019.10.002] [Medline: 31672660]
18. Amorim J, Ventura AC. Co-created decision-making: from co-production to value co-creation in health care. J Med Access 2023 Jun 6;7:27550834231177503. [doi: 10.1177/27550834231177503] [Medline: 37323851]
19. Kenis I, Kinnaer LM, Delombaerde C, et al. A self-directed co-creation process for developing a care pathway for patients on oral anticancer therapy: a qualitative process evaluation. Eur J Oncol Nurs 2023 Aug;65:102353. [doi: 10.1016/j.ejon.2023.102353] [Medline: 37321129]
20. Kinnaer LM, Van de Vyver M, Kenis I, Decoene E, Foulon V, Van Hecke A. A qualitative evaluation of the process of creating and implementing an interprofessional care pathway for patients treated with oral anticancer drugs. Eur J Oncol Nurs 2022 Dec;61:102218. [doi: 10.1016/j.ejon.2022.102218] [Medline: 36395703]
21. Murray AL, Alpern E, Lavelle J, Mollen C. Clinical pathway effectiveness: febrile young infant clinical pathway in a pediatric emergency department. Pediatr Emerg Care 2017 Sep;33(9):e33-e37. [doi: 10.1097/PEC.0000000000000960] [Medline: 28072664]




XSL•FO
RenderX

22. Janamian T, Crossland L, Jackson CL. Embracing value co-creation in primary care services research: a framework for success. Med J Aust 2016 Apr 18;204(7 Suppl):S5-S11. [doi: 10.5694/mja16.00112] [Medline: 27078794]
23. Grindell C, Coates E, Croot L, O'Cathain A. The use of co-production, co-design and co-creation to mobilise knowledge in the management of health conditions: a systematic review. BMC Health Serv Res 2022 Jul 7;22(1):877. [doi: 10.1186/s12913-022-08079-y] [Medline: 35799251]
24. Millard JW, Agha YH, Srinivasan S, Assi M. The effect of a clinical pathway on reducing the rate of healthcare-onset Clostridioides difficile. Kans J Med 2020 Oct;13:260-264. [doi: 10.17161/kjm.vol13.14762] [Medline: 33173561]
25. Singh SB, Shelton AU, Greenberg B, Starner TD. Implementation of cystic fibrosis clinical pathways improved physician adherence to care guidelines. Pediatr Pulmonol 2017 Feb;52(2):175-181. [doi: 10.1002/ppul.23635] [Medline: 27797455]
26. Mater W, Aldwairi M, Ibrahim R. Enhanced teamwork communication model for electronic clinical pathways in healthcare. Open Bioinform 2018 Jul 31;11(1):140-163. [doi: 10.2174/1875036201811010140]
27. Montesanti S, Pritchard C, Green L, Berg S, Mallinson S, Birdsell J. Implementing primary health care teams and integrated care in Alberta, Canada. Health Reform Obs 2022 Apr 1;10(1). [doi: 10.13162/hro-ors.v10i1.4680]
28. Kain N, Ashworth N, Hernandez-Ceron N, Hamayeli-Mehrabani H, Hurava I, Kumar K. Predictors of high-performing family medicine clinics: prospective cohort study in Alberta. Can Fam Physician 2026 Jun;72(6):400-407. [doi: 10.46747/cfp.7206400] [Medline: 42285725]
29. Wangler J, Jansky M. Primary care involvement in clinical research - prerequisites, motivators, and barriers: results from a study series. Arch Public Health 2024 Mar 20;82(1):41. [doi: 10.1186/s13690-024-01272-x] [Medline: 38504310]
30. Gil Conde M, Penedo C, Barcelos FF, et al. Strategies for research capacity building by family physicians in primary care: a scoping review. J Prim Health Care 2026 Sep 24;18(3):256-269. [doi: 10.1071/HC25042] [Medline: 41134006]
31. Malmqvist J, Hellberg K, Möllås G, Rose R, Shevlin M. Conducting the pilot study: a neglected part of the research process? Methodological findings supporting the importance of piloting in qualitative research studies. Int J Qual Methods 2019 Sep 25;18:1609406919878341. [doi: 10.1177/1609406919878341]


## Abbreviations

**AHS:** Alberta Health Services
**CDI:** *Clostridioides difficile* infection
**FMT:** fecal microbiota transplantation
**PCP:** primary care provider
**PCR:** polymerase chain reaction
**rCDI:** recurrent *Clostridioides difficile* infection


*Edited by M Balcarras; submitted 18.Feb.2026; peer-reviewed by A Girma, CA Warren, P Shingru; revised version received 14.Aug.2026; accepted 16.Sep.2026; published 30.Sep.2026.*

*Please cite as:*
*Patel RK, Franz R, Tate J, Coppens J, Do Rego M, Smith C, Ovadje L, Wong K, Saxinger L, Kao D*
*Codeveloping a Clinical Pathway for Recurrent Clostridioides difficile Infection With Primary Care Providers: User-Centered Design Study*
*Interact J Med Res 2026;15:e93032*
*URL: https://www.i-jmr.org/2026/1/e93032*
*doi:10.2196/93032*



© Ritika Kamlesh Patel, Rose Franz, Jordan Tate, Jennifer Coppens, Matthew Do Rego, Christopher Smith, Lauretta Ovadje, Karen Wong, Lynora Saxinger, Dina Kao. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 30.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.




XSL•FO
RenderX

# Impact of mHealth-Augmented Social Support on Health Care Use Among Patients With Diabetes: Secondary Analysis of the TExT-MED+FANS Trial

Danielle Hazime[1,2,3*], BS, MHA; Liza Raffi[4*], BS, MD; Elizabeth Burner[5], BS, MD, PhD

[1]Rush Medical College of Rush University Medical Center, Chicago, IL, United States

[2]Cancer Research Center for Health Equity, Cedars-Sinai Medical Center, 700 N San Vicente Boulevard, Suite G-500, Los Angeles, CA, United States

[3]Sol Price School of Public Policy, University of Southern California, Los Angeles, CA, United States

[4]Department of Internal Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, United States

[5]Department of Emergency Medicine, Keck School of Medicine of USC, Los Angeles, CA, United States

[*]these authors contributed equally

**Corresponding Author:**
Danielle Hazime, BS, MHA
Rush Medical College of Rush University Medical Center, Chicago, IL, United States

## Abstract

**Background:** The rising cost of unscheduled acute health care, particularly for emergency department (ED) visits, poses significant financial burdens. In 2021, aggregate costs for treat-and-release ED visits in the United States accounted for an estimated US $80 billion, while the total annual cost of diabetes was US $412.9 billion in 2022—representing about 1 in every 4 health care dollars, 61% of which are directly linked to diabetes.

**Objective:** This study explores the impact of a mobile health (mHealth) intervention with augmented social support delivered via guided SMS text messaging on health care use among patients with diabetes through a secondary analysis of the TExT-MED+FANS (Trial to Examine Text-Messaging in Emergency Patients With Diabetes + Family and Friends Network Support) randomized controlled trial.

**Methods:** The trial involved 173 participants randomized into either a FANS mHealth-augmented social support or an active control group that received the same support curriculum via mailed pamphlet; “augmented” social support refers to recruiting both individuals with diabetes and a designated family member or friend (“supporter”) to participate. Supporters in the FANS arm received structured SMS guidance on how to assist the participant with diabetes in managing their condition. Health care use outcomes, including ED visits, hospitalizations, and clinic visits, were compared between groups during and after the intervention period using linear regression models on change in health care visits in the last year, with subgroup analysis by participant sex and supporter relationship.

**Results:** Results showed significant reductions in acute unscheduled care visits for both groups during and after the intervention, with the FANS group experiencing a reduction of 1.04 visits during the intervention and 1.10 visits after the intervention, while the mailed pamphlet group had reductions of 1.47 and 1.53 visits, respectively (both $P<.001$). Clinic visits increased by 1.78 during the intervention phase ($P=.01$) but were not sustained postintervention. Hospitalizations modestly decreased in both groups, but the decrease was statistically significant only for the mailed pamphlet group ($P=.002$). Sex and supporter relationship differences were observed, with females supported by spouses in the active control group showing the largest decrease in unscheduled care visits.

**Conclusions:** The findings suggest that mHealth interventions combined with structured social support can improve diabetes management and reduce health care costs. Although these conclusions are specific to this study, they align with prior research demonstrating the benefits of social support and mHealth interventions. These findings may inform future programs, including the design of low-cost, scalable interventions in resource-limited settings.

**Trial Registration:** ClinicalTrials.gov NCT03178773; https://www.clinicaltrials.gov/study/NCT03178773

**International Registered Report Identifier (IRRID):** RR2-10.1016/j.cct.2019.03.003

***(Interact J Med Res 2026;15:e65113)*** doi:10.2196/65113

**KEYWORDS**

diabetes; health care use; mHealth; mobile health; social support; health care utilization



XSL•FO
RenderX

## Introduction

The cost of unscheduled, acute health care, particularly for emergency department (ED) care, continues to rise, with the strain particularly evident among patients with diabetes. In 2021, aggregate costs for treat-and-release ED visits in the United States accounted for an estimated US $80 billion in costs, with the total annual cost of diabetes being US $412.9 billion in 2022, representing about 1 in every 4 health care dollars—61% of which are directly linked to diabetes [1,2]. Patients with diabetes may need frequent unscheduled care due to complicated care needs, difficulty accessing care, and lack of self-management skills that result in clinical decompensation [3,4]. The US health care system often faces heavy financial burdens resulting from unscheduled visits [2,5]. In addition, high medical bills and unpredicted lost productivity ultimately pose significant expenses to patients themselves [6]. This is compounded by the broader societal impact, where increased health care spending can affect economic stability. Finding strategies to minimize these costs is important to creating a more sustainable health care ecosystem.

Within the context of such challenges, mobile health (mHealth) interventions offer an approach to managing chronic diseases like diabetes, providing tools such as glucose monitoring, medication reminders, dietary tracking, and personalized coaching [7,8]. These platforms aim to reduce health care costs, improve accessibility, and enhance patient medication taking, engagement, and involvement [9-11]. Examining the impact of mHealth interventions on health care use is crucial for evaluating their feasibility and scalability across various health care settings and populations, as they can greatly reduce direct costs [12,13].

Social support plays an important role in diabetes management, with prior studies showing that family and friend involvement can improve medication taking, self-management behaviors, and glycemic stability [14-16]. Recent work has also examined mHealth-directed social support, suggesting that combining digital health tools with structured support may further enhance patient engagement and outcomes [16]. However, few studies have directly compared mHealth-augmented support with traditional in-person support or print-based approaches. Such comparisons are essential to inform comprehensive diabetes management strategies that effectively leverage both technology and social networks to improve patient care and reduce costs associated with unscheduled health care use.

To understand the impact of mHealth integrated social support on health care use, this study offers a detailed analysis of health care use trends as a secondary analysis of TExT-MEDS+FANS (Trial to Examine Text-Messaging in Emergency Patients With Diabetes + Family and Friends Network Support), a randomized controlled trial of mHealth-augmented social support (FANS) compared to standard social support (mailed pamphlet) added to a patient-focused mHealth curriculum for patients with diabetes. The primary research objectives were to assess the association between mHealth-delivered social support and its influence on health care–seeking behaviors, including the frequency and nature of health care use. We also examined whether the character of the existing support associations prior to the intervention modified the intervention effect.

## Methods

### Research Methodology

TExT-MED+FANS was a 12-month comparative effectiveness randomized controlled trial that investigated the impact of two exposures: (1) mHealth-augmented social support (FANS) and (2) standard social support (mailed pamphlet) on patient health care use outcomes. Participants were enrolled from June 2017 to December 2019. This study represents a secondary analysis of the original randomized trial data, focused on differences in health care periods and subgroup characteristics not examined in the primary publication. For further information on the methods used in TExT-MED+FANS and the specific mHealth program under study, please refer to the protocol study by Burner et al [17].

### Ethical Considerations

The study received approval from the University of Southern California Health Sciences Institutional Review Board (HS-17 - 00406). All participants provided written informed consent. Data were stored in an encrypted dataset and were not deidentified to allow linkage to electronic medical record (EMR) information. Consent procedures were conducted in English or Spanish, depending on patient preference. Patients and supporters were provided with compensation ranging from US $20 to US $280, depending on the number of study visits completed [16].

### Study Population

#### *Study Protocol: Screening*

Patients were eligible to participate if they had either type 1 or type 2 diabetes with elevated glycated hemoglobin ($A_{1c}$) levels ($A_{1c} \geq 8.5$), were aged 18 years or older, possessed the capability to send and receive text messages, could identify a supporter, and were able to provide informed consent.

#### *Enrollment of Patient*

Trained research assistants used convenience sampling by approaching eligible patients with diagnosed diabetes identified through their EMR during routine ED visits, regardless of the nature of the patient's visit. No public advertisement or broader outreach strategy was used. Patients provided informed consent and completed an intake interview in their preferred language (English or Spanish). They were then registered in the SMS text-messaging platform to access the TExT-MED patient curriculum. While in the ED, patients nominated a family member or close friend to serve as their supporter for the study, and research assistants collected contact information for up to 3 supporters.

#### *Enrollment of Support Person*

Support person enrollment took place concurrently in the ED or through telephone communication within a 2-week window following patient enrollment. Supporters completed the enrollment process by providing informed consent and verifying their ability to send and receive text messages.



XSL•FO
RenderX

### *Randomization*

Following supporter enrollment, patient-supporter dyads were randomized to either mHealth-augmented social support through the FANS program or standard social support delivered via mailed pamphlets. If the dyad was randomized to the FANS arm, research assistants registered the family member in the SMS text-messaging platform to receive the mHealth-augmented support curriculum. A computer-generated simple randomization sequence was used, with allocation concealed via sealed envelopes until assignment.

### *Treatment Groups: Exposures*

The main exposure variable was the type of social support intervention: mHealth-augmented support (FANS) versus standard support (mailed pamphlet). Every patient received the SMS text message–based patient curriculum (TExT-MED) regardless of their group assignment. This included automated, static (noninteractive) text messages delivered daily for 6 months. The content included education on diabetes self-management, encouragement, and reminders. Messages were not personalized. Supporters in the intervention group received a text message–based curriculum (FANS) guiding them to provide mHealth-augmented social support to their loved ones. These messages were also static and noninteractive, delivered twice weekly. Supporters in the active control group received standard social support through the same material and curriculum distributed via a mailed pamphlet. No additional support or follow-up beyond the pamphlet was provided to the control group supporters.

## Measures: Outcomes and Assessments

The primary outcome was health care use, assessed through EMR review. Outcome variables that were included are given below:

1. Acute unscheduled care, defined as any visit to the Medical Center ED or urgent care.
2. Hospitalizations, defined as any inpatient stay recorded in the Los Angeles County Department of Health Services system.
3. Scheduled clinic visits, defined as fulfilled clinic appointments with a primary care provider within the Los Angeles County Department of Health Services.

All calculations underwent scrutiny by 2 reviewers, with any discrepancies resolved through consensus.

## Measurement Phases

The following measurement phases were used to assess changes in health care use over time:

- Prior: 6 months prior to enrollment.
- Active: from the intervention start date to the intervention end date (a 6-month period).
- After intervention: end of intervention date to 6 months after intervention. Participants did not receive SMS text messages during this period.

At the time of enrollment, patients self-reported their sex, race or ethnicity, primary language, duration of diabetes, and access to primary care. Age and insurance status were extracted from the EMR. Insurance status was classified as means-tested (eg, Medicaid or other government insurance programs), non–means-tested (eg, Medicare or private insurance), or uninsured (no recorded insurance coverage at the time of the visit). The supporter's sex was reported by the supporter, and for analysis purposes, supporters identified as cohabitating significant others by the patient were considered spouses. Supporters other than spouses were categorized as nonspouse supporters, encompassing various relationships such as siblings, friends, or children.

## Data Analysis

The intervention group assignment was otherwise examined with an intention-to-treat approach, regardless of patient or supporter dropping out of the TExT-MED or FANS text messages. Analysis was completed with STATA (version 17.0; StataCorp LLC). Data were examined with histograms and did not violate assumptions of heteroscedasticity. Baseline characteristics were examined descriptively as a whole and by intervention group (Table 1).

In our main effects analysis, we used linear regression models with outcomes of change scores of health care visits from (1) baseline to the end of intervention phases, and (2) from the end of intervention to the end of maintenance phase, examining differences in health care use by the FANS augmented versus traditional mailed pamphlet groups (Table 2). The main independent variable was intervention group assignment. No other independent variables were included in the main effects model. Differences between groups were estimated with the contrast and margins function. We individually examined unscheduled acute care visits (ED and urgent care visits), hospitalizations, and scheduled care (clinic) visits. For subgroup analysis, we used an interaction term for sex and supporter relationship to examine the moderating effect of supporter relationship and sex with the same procedures (Table 3).



XSL•FO
RenderX

**Table .** Demographic data (N=165).

| Characteristics | Active control (n=86) | FANS[a] (n=79) | Combined |
|---|---|---|---|
| Age (y), mean (SD) | 48.25 (11) | 46.88 (10) | 47.59 (10.5) |
| Sex (male), n (%) | 37 (43) | 45 (57) | 82 (49.7) |
| Spouse as supporter, n (%) | 37 (43) | 29 (37) | 66 (40) |
| Race or ethnicity, n (%) | | | |
| Asian | 2 (2) | 0 (0) | 2 (1.2) |
| Black | 3 (3) | 6 (8) | 9 (5.5) |
| Hispanic | 81 (94) | 71 (90) | 152 (92.1) |
| Non-Hispanic White | 0 (0) | 2 (3) | 2 (1.2) |
| Primary language (Spanish), n (%) | 64 (74) | 51 (65) | 115 (69.7) |
| Insurance type, n (%) | | | |
| Means-based | 82 (95) | 71 (90) | 153 (92.7) |
| Non–means-based | 2 (2) | 5 (6) | 7 (4.2) |
| Uninsured | 2 (2) | 3 (4) | 5 (3) |
| Self-reported at least 1 primary care visit(s) in 6 months prior to study, n (%) | 64 (74) | 63 (80) | 127 (77) |
| Years with diabetes, mean (SD) | 11.65 (9) | 11.10 (9) | 11.39 (8.7) |
| Did not complete full curriculum, n (%) | 6 (7) | 4 (5) | 10 (6.1) |

[a]FANS: Family and Friends Network Support.



XSL•FO
RenderX

**Table .** Main outcomes (N=165)[a].

| Outcome | Active control (mailed pamphlet; n=86) | | FANS[b] (mHealth-augmented social support; n=79) | | Between-group difference in change score (active control—FANS) | | Combined total study population | |
|---|---|---|---|---|---|---|---|---|
| | Mean (95% CI) | *P* value | Mean (95% CI) | *P* value | Mean (95% CI) | *P* value | Mean (95% CI) | *P* value |
| Δ-Acute unscheduled care visits intervention phase versus patient's baseline | −1.47 (−1.87 to −1.06) | <.001 | −1.04 (−1.41 to −0.66) | <.001 | −0.43 (−0.98 to 0.12) | .13 | −1.26 (−1.54 to −0.99) | <.001 |
| Δ-Acute unscheduled care visits postintervention phase versus patient's baseline | −1.53 (−1.93 to −1.14) | <.001 | −1.10 (−1.52 to −0.69) | <.001 | −0.43 (−1.00 to 0.13) | .13 | −1.33 (−1.61 to −1.04) | <.001 |
| Δ-Clinic visits intervention phase versus patient's baseline | 1.00 (0.13 to 1.80) | .02 | 1.78 (0.40 to 3.17) | .01 | −0.82 (−2.39 to 0.75) | .31 | 1.36 (0.57 to 2.14) | <.001 |
| Δ-Clinic visits postintervention phase versus patient's baseline | −0.14 (−1.20 to 0.92) | .79 | 0.48 (−0.81 to 1.77) | .46 | −0.62 (−2.26 to 1.02) | .46 | 0.16 (0.81 to 1.00) | .70 |
| Δ-Hospitalizations (intervention phase vs patient's baseline) | −0.24 (−0.40 to 00.09) | .002 | −0.18 (−0.37 to 0.01) | .07 | −0.07 (−0.31 to 0.17) | .58 | −0.21 (−0.33 to −0.09) | <.001 |
| Δ-Hospitalizations (postintervention phase vs patient's baseline | −0.22 (−0.38 to −0.06) | .01 | −0.20 (−0.44 to 0.04) | .01 | −0.02 (−0.30 to 0.26) | .90 | −0.21 (−0.35 to −0.07) | .003 |

[a]Values represent the mean change in health care use (with 95% CIs) from baseline to each study period for the intervention and control groups. *P* values reflect the significance of within-group changes, while the between-group difference column represents the difference in change scores between groups.

[b]FANS: Family and Friends Network Support.



XSL•FO
RenderX

**Table .** Sex and supporter type subanalysis[a].

| | Female with spouse supporters, mean (95% CI; *P* value) | | Females with nonspouse supporters, mean (95% CI; *P* value) | | Male with spouse supporters, mean (95% CI; *P* value) | | Males with nonspouse supporters, mean (95% CI; *P* value) | |
|---|---|---|---|---|---|---|---|---|
| | Active control | FANS[b] | Active control | FANS | Active control | FANS | Active control | FANS |
| Acute unscheduled care | | | | | | | | |
| Δ-6-month | −2.41 (−3.24 to −1.59; *P*<.001) | 0.63 (0.58 to 1.83; *P*=.31) | −1.25 (−1.85 to −0.65; *P*<.001) | –1.58 (-2.25 to -0.91; *P*<.001) | –1.05 (-1.81 to -0.29; *P*=.007) | –1.19 (-1.93 to -0.45; *P*=.002) | –1.41 (–2.24 to –0.59; *P*<.001) | –0.88 (–1.57 to –0.18; *P*=.01) |
| Δ-12-month | −2.76 (−3.61 to −1.92; *P*<.001) | 0.38 (−0.86 to 1.61; *P*=0.55) | −1.25 (−1.87 to −0.63; *P*<.001) | −1.42 (−2.11 to −0.74; *P*<.001) | −0.9 (−1.68 to −0.12; *P*=.02) | −1.00 (−1.76 to −0.24; *P*=0.01) | −1.59 (–2.44 to –0.74; *P*<.001) | –1.33 (–2.05 to –0.62; *P*<.001) |
| Clinic visits | | | | | | | | |
| Δ-6-month | 2.47 (0.05 to 4.89; *P*=.05) | 4.00 (0.47 to 7.53) (*P*=.03) | 1.50 (−0.26 to 3.3) (*P=.10)* | 0.04 (−1.92 to 1.99) (*P*=.97) | −0.35 (−2.58 to 1.88) (*P*=.76) | 1.48 (−0.70 to 3.65) (*P*=.18) | 0.00 (−2.42 to 2.42) (*P*=1.00) | 3.21 (1.17 to 5.24) (*P*=.002) |
| Δ-12-month | 1.12 (−1.40 to 3.63; *P*=.38) | 4.75 (1.08 to 8.42; *P*=.011) | 0.03 (−1.80 to 1.87; *P*=.97) | −0.62 (−2.65 to 1.42; *P*=.55) | −1.55 (−3.87 to 0.77; *P*=.19) | −1.05 (−3.31 to 1.22; *P*=.36) | −0.06 (−2.57 to 2.46; *P*=.96) | 1.58 (−0.53 to 3.70; *P*=.14) |
| Hospitalizations | | | | | | | | |
| Δ-6-month | −0.18 (−0.55 to 0.20; *P*=.36) | −0.38 (−0.92 to 0.17; *P*=.18) | −0.28 (−0.56 to −0.01; *P*=.05) | 0.04 (−0.27 to 0.34; *P*=.80) | −0.25 (−0.60 to 0.10; *P*=.16) | −0.24 (−0.58 to 0.10; *P*=.17) | −0.24 (−0.61 to 0.14; *P*=.22) | −0.29 (−0.61 to 0.03; *P*=.07) |
| Δ-12-month | −0.24 (−0.68 to 0.21; *P*=.30) | -0.25 (−0.90 to 0.40; *P*=.45) | −0.19 (−0.51 to 0.14; *P*=.26) | −0.12 (−0.48 to 0.24; *P*=.53) | −0.3 (−0.71 to 0.11; *P*=.15) | −0.05 (−0.45 to 0.35; *P*=.82) | −0.18 (−0.62 to 0.27; *P*=.44) | −0.42 (−0.79 to −0.04; *P*=.03) |

[a]Values represent the mean change in health care use (with 95% CIs) by sex and supporter relationship for each study period. *P* values indicate the significance of within-group changes.

[b]FANS: Family and Friends Network Support.

## Results

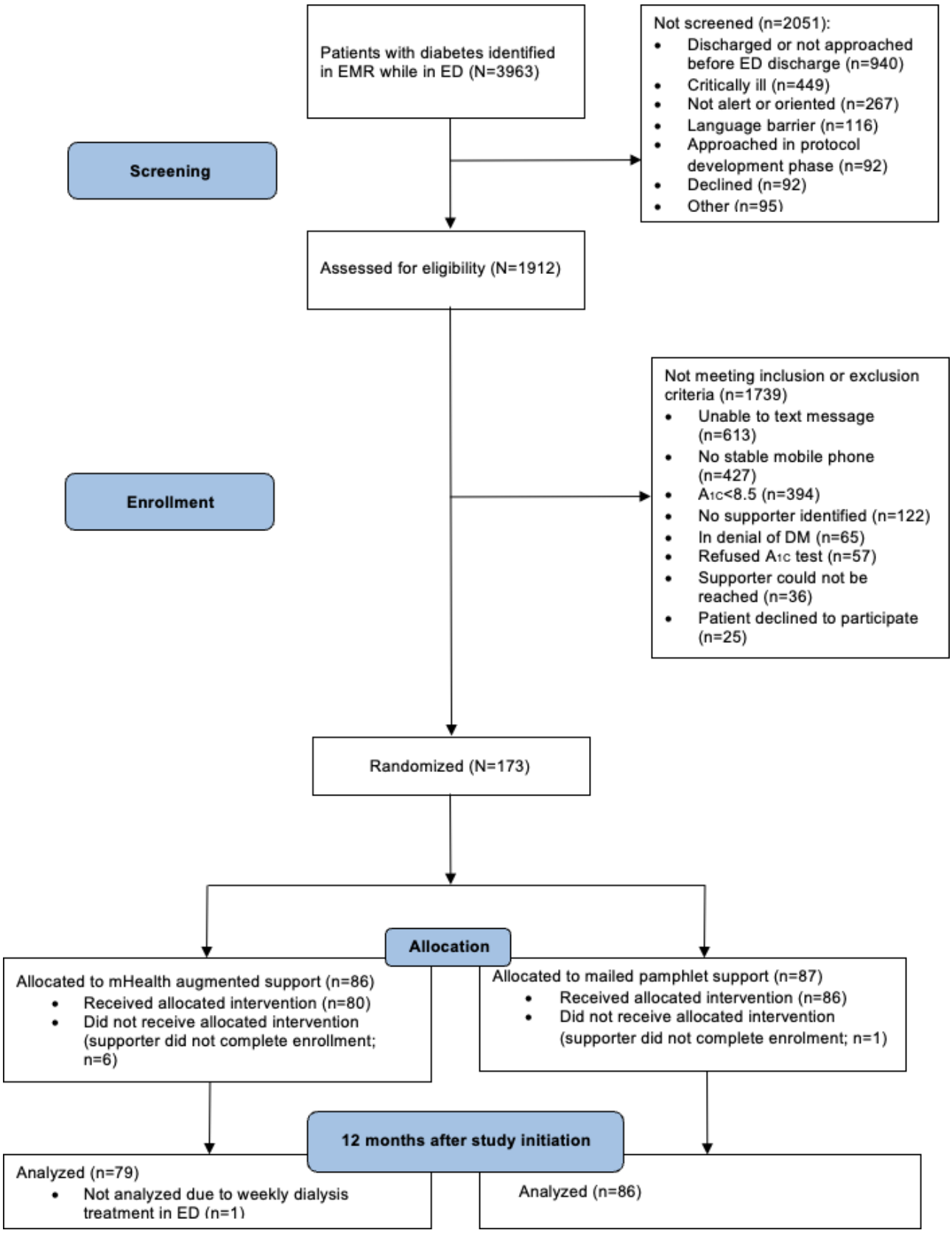
In this study, a total of 3963 patients with diabetes were identified from electronic health records while in the ED (Figure 1). Of these, 12051 patients were not screened for various reasons, including critical illness, discharge or nonapproach before ED discharge, lack of alertness and orientation, language barriers, and other reasons. Among the screened individuals, 1739 were excluded due to inability to send SMS text messages, lack of a stable mobile phone, $A_{1c}$ <8.5%, absence of a supportive individual, and other reasons (n=183; Figure 1). Ultimately, 173 patients were randomized into the study and allocated either to mHealth-augmented support (86/173, 49.7%) or standard social support (mailed pamphlet; 87/173, 50.3%). One participant in the mHealth-augmented group was excluded from this analysis because they were undergoing thrice-weekly dialysis treatment in the ED due to insurance limitations during the study period. The final analytic population was composed of 165 participants randomized into 2 intervention groups: FANS with 79 participants and a mailed pamphlet with 86 participants.

The average age of a patient was 47.59 (10.5) years, and males comprised approximately half of the study population, with a slightly higher proportion in the FANS group (45/79, 57%) compared to the mailed pamphlet group (37/86, 43%). The majority of participants self-identified as Latino or Hispanic (152/165, 92.1%), with small representations from Non-Latino or Hispanic White (2/165, 1.2%), Asian (2/165, 1.2%), and Black (9/165, 5.5%) backgrounds. A substantial proportion of participants primarily spoke Spanish (115/165, 69.7%). Insurance coverage was predominantly means-based (Medicaid or other means-based government insurance programs; 153/165, 92.7%), with a few participants either non–means-based insured (Medicare or private insurance) (7/165, 4.2%) or uninsured (5/165, 3%). Most participants had visited a primary care provider within the previous 6 months (127/165, 77%). The average duration of diabetes among the participants was 11.39 (9.7) years, indicating a long-standing condition of diabetes within the cohort (Table 1).



XSL•FO
RenderX

**Figure 1.** Participant flow diagram. DM: diabetes mellitus; ED: emergency department; EMR: electronic medical record; FANS: Family and Friends Network Support; mHealth: mobile health.




The results from the study indicated significant changes in health care use across different types of care visits during the intervention and postintervention phases (Table 2). During the intervention phase, there was a notable reduction in acute unscheduled care visits across both intervention groups. The FANS group experienced an average decrease of 1.04 visits ($P$<.001), while the mailed pamphlet group saw a larger average decrease of 1.47 visits ($P$<.001). This trend was sustained after intervention, with reductions of 1.10 and 1.53 visits, respectively, for FANS and mailed pamphlet groups, both



XSL•FO
RenderX

statistically significant ($P$<.001). The difference between the FANS and mailed pamphlet groups was not statistically significant in either period.

Clinic visits increased during the intervention phase, with the FANS group showing a mean increase of 1.78 visits ($P$=.01) and the mailed pamphlet group an increase of 1.00 visits ($P$=.02). However, this increase was not sustained after intervention, with the changes becoming statistically insignificant and the direction of change mixed. There was a difference of 0.62 visits between the intervention groups during the postintervention period; however, this was not statistically significant.

Hospitalization rates showed a modest decrease during the intervention phase with reductions of 0.24 visits in the mailed pamphlet group and 0.18 visits in the FANS group, with the former being statistically significant ($P$=.002). This decreasing trend in hospitalizations continued into the postintervention phase with similar reductions. There was a difference of −0.07 between the intervention groups, though this was not statistically significant.

During the 6-month interval, females with spouse supporters in the mailed pamphlet group experienced a significant decrease in acute unscheduled care visits, showing a reduction of 2.41 visits ($P$<.001). In contrast, males with spouse supporters saw a decrease of 1.19 visits ($P$=.002) in the FANS group and a similar decrease in the mailed pamphlet group. Females with nonspouse supporters and males with nonspouse supporters also showed significant decreases, though the effect was strongest in those with spouse supporters. Clinic visits increased significantly during the 6-month interval for both females and males with spouse supporters in the FANS group, with increases of 4 visits ($P$=.03) and 3.21 visits ($P$=.002), respectively.

## Discussion

### Principal Findings

In this study, we evaluated the impact of a mHealth-augmented social support intervention (FANS) compared to an active control (mailed pamphlet) on health care use among patients with diabetes, analyzing data from 165 participants to assess changes in acute unscheduled care visits, clinic visits, and hospitalizations. The main findings indicate a significant reduction in acute unscheduled care visits and hospitalizations for both intervention and active control groups, with this effect maintained after intervention. Moreover, an increase in clinic visits was seen during both the intervention and postintervention phases. In addition, we found that the type of supporter, particularly whether or not the supporter was a spouse, played a role in study participants' use of the mentioned health services.

The reduction in acute unscheduled care visits aligns with existing literature supporting the effectiveness of mHealth interventions and educational pamphlets in improving patient self-management [18]. Previous studies have shown that providing structured support and education to patients with chronic diseases can lead to better health outcomes [19]. Specifically for patients with diabetes, these interventions have been linked to improved diabetes care, glycemic stability, and reduced $A_{1c}$ levels [20]. Moreover, the integration of mHealth technologies has facilitated more consistent monitoring and timely interventions, enhancing overall disease management. Such approaches not only empower patients but also foster a proactive engagement in their health care journey, ultimately reducing the burden on acute care services [21].

There is a notable gap in the literature regarding health care use following mHealth interventions, particularly concerning ED visits, hospitalizations, and clinic visits for patients with diabetes. While extensive data exists on the impact of such interventions on other chronic conditions, like chronic obstructive pulmonary disease (COPD), similar comprehensive studies on diabetes-focused mHealth interventions are lacking. Yang et al [22] demonstrated that an mHealth curriculum for patients with COPD significantly reduced the risk of hospitalizations, although it did not yield a substantial difference in the average length of stay. In this study focusing on diabetes, a marked reduction in the frequency of hospitalizations was observed both during and after the intervention; however, we did not examine the length of stay. While both conditions benefit from mHealth interventions, the nature and progression of diabetes and COPD may present unique challenges and opportunities for reducing healthcare utilization, emphasizing the need for more tailored research in diabetes-specific programs.

The subgroup analysis revealed that while acute unscheduled care visits decreased significantly across most supporter groups, females with spouse supporters in the FANS group showed no significant change. The type of supporter may influence the intervention's effectiveness, aligning with prior work indicating that spousal support by sex can vary widely in its impact on health behaviors [23]. These patterns underscore the complex role of social support dynamics, as highlighted in previous studies which have shown that the quality and type of support can differentially impact health outcomes [24,25]. Specifically, research has emphasized the effectiveness of spousal support in chronic disease management, including diabetes, due to the consistent and close nature of the support provided [26]. Clinic visits increased significantly for females with spouse supporters during the intervention phase in both groups, which aligns with studies indicating that females are generally more proactive in seeking preventive care [27-29]. In this study, this increase was only sustained in the FANS group, suggesting the potential for digital health tools to maintain engagement with health care services. This study's findings align with these insights, suggesting that while mHealth interventions can be broadly beneficial, their impact may be moderated by the type of support relationship involved.

The findings of this study have substantial implications for public health and policy. The reduction in acute unscheduled care visits among patients with diabetes through mHealth-augmented social support interventions highlights the potential of such technologies in enhancing chronic disease management. Policy makers should consider integrating mHealth solutions into standard care protocols, particularly for populations with limited access to care, to improve health care accessibility and outcomes [20,30,31]. In addition, investing in



XSL•FO
RenderX

mHealth platforms could alleviate the financial burden on health care systems by reducing the frequency of costly acute unscheduled care visits and hospitalizations [32,33]. Public health initiatives should also focus on promoting digital literacy and access to mHealth technologies, ensuring that communities that have been historically marginalized can fully benefit from these advancements. By incorporating mHealth interventions into broader health care strategies, a more sustainable and equitable health care system can be created that better addresses the needs of patients with chronic conditions like diabetes. Future research and policy initiatives should also consider incorporating structured family or supporter-based engagement as a mechanism to enhance intervention effectiveness and long-term adherence.

This study has several limitations that must be considered when interpreting the findings. First, the small sample size and the secondary analysis nature of the study may limit the internal validity of the results. Using secondary data introduces inherent limitations, including potential bias related to how the original study was designed and executed. Since this analysis was not prespecified in the original study protocol, there may be risks of selective reporting or post hoc interpretation. In addition, certain variables of interest may not have been collected or controlled for in the original trial design, potentially limiting the scope of inferences. Future research with larger, more diverse populations is needed to confirm these findings. Second, acute unscheduled care visits in this study included both urgent care and ED visits, due to the structure of the Los Angeles County Department of Health Service Electronic Health Record System. While urgent care visits are less costly than ED visits, they still are inferior to scheduled clinic care and are not the desired health care use pattern [34,35]. Third, the sample was predominantly composed of individuals from Hispanic or Latino and low-socioeconomic backgrounds who used means-based insurance. These findings may not be generalizable to populations with different demographic and socioeconomic characteristics. Despite these limitations, the study provides valuable insights into the potential benefits of mHealth interventions. Addressing limitations and exploring tailored approaches can enhance the efficacy and reach of such interventions.

## Conclusions

This study contributes to the growing body of evidence examining the use of mHealth tools to support chronic disease management, reduce the burden on health care systems, and support the effectiveness of structured support systems and educational interventions in enhancing patient self-management and reducing health care use. mHealth-augmented social support interventions, such as the FANS program, may be able to reduce acute unscheduled care visits and hospitalizations among patients with diabetes. Though more work in this area is needed to identify who will most benefit, we saw indications that these interventions, particularly when involving spouse supporters, can lead to meaningful improvements in health care outcomes and engagement with preventive care services. The significant reductions in unscheduled care visits suggest that mHealth-augmented social support can effectively improve diabetes management, particularly for disproportionately affected groups. These findings underscore the importance of incorporating digital health solutions into chronic disease management social support strategies to improve health care outcomes and reduce acute care reliance.

## Acknowledgments

The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content of this manuscript.

## Funding

Funding for this trial was provided by National Institute of Diabetes and Digestive and Kidney Diseases Award 5K23DK106538.

## Data Availability

The datasets generated or analyzed during this study are not publicly available due to institutional review board restrictions but are available from the primary investigator upon reasonable request.

## Authors' Contributions

Conceptualization: LR, DH, EB
Data curation: EB
Formal analysis: EB
Funding acquisition: EB
Investigation: DH, LR, EB
Methodology: LR, EB
Project administration: EB
Supervision: EB
Visualization: DH
Writing – original draft: DH, LR, EB
Writing – review & editing: DH, LR, EB



XSL•FO
RenderX

## Conflicts of Interest

EB has received compensation as a consultant and on advisory boards from Agile Health on unrelated SMS-based interventions for patient self-care.

## References


1. Roemer M. Costs of treat-and-release emergency department visits in the United States, 2021. : Agency for Healthcare Research and Quality (AHRQ); 2024 URL: https://hcup-us.ahrq.gov/reports/statbriefs/sb311-ED-visit-costs-2021.pdf [accessed 2025-11-20]
2. Parker ED, Lin J, Mahoney T, et al. Economic costs of diabetes in the U.S. in 2022. Diabetes Care 2024 Jan 1;47(1):26-43. [doi: 10.2337/dci23-0085] [Medline: 37909353]
3. Benoit SR, Hora I, Pasquel FJ, Gregg EW, Albright AL, Imperatore G. Trends in emergency department visits and inpatient admissions for hyperglycemic crises in adults with diabetes in the U.S., 2006-2015. Diabetes Care 2020 May;43(5):1057-1064. [doi: 10.2337/dc19-2449] [Medline: 32161050]
4. Mayberry LS, Mulvaney SA, Johnson KB, Osborn CY. The MEssaging for Diabetes intervention reduced barriers to medication adherence among low-income, diverse adults with type 2. J Diabetes Sci Technol 2017 Jan;11(1):92-99. [doi: 10.1177/1932296816668374] [Medline: 27595710]
5. American Diabetes Association. Economic costs of diabetes in the U.S. in 2017. Diabetes Care 2018 May;41(5):917-928. [doi: 10.2337/dci18-0007] [Medline: 29567642]
6. Rocha JVM, Marques AP, Moita B, Santana R. Direct and lost productivity costs associated with avoidable hospital admissions. BMC Health Serv Res 2020 Mar 13;20(1):210. [doi: 10.1186/s12913-020-5071-4] [Medline: 32164697]
7. Lee JL, Kim Y. Evaluation of mobile applications for patients with diabetes mellitus: a scoping review. Healthcare (Basel) 2024 Jan 31;12(3):368. [doi: 10.3390/healthcare12030368] [Medline: 38338253]
8. Sun C, Malcolm JC, Wong B, Shorr R, Doyle MA. Improving glycemic control in adults and children with type 1 diabetes with the use of smartphone-based mobile applications: a systematic review. Can J Diabetes 2019 Feb;43(1):51-58. [doi: 10.1016/j.jcjd.2018.03.010] [Medline: 30026048]
9. Eberle C, Löhnert M, Stichling S. Effectiveness of disease-specific mHealth apps in patients with diabetes mellitus: scoping review. JMIR Mhealth Uhealth 2021 Feb 15;9(2):e23477. [doi: 10.2196/23477] [Medline: 33587045]
10. Hou C, Xu Q, Diao S, Hewitt J, Li J, Carter B. Mobile phone applications and self-management of diabetes: a systematic review with meta-analysis, meta-regression of 21 randomized trials and GRADE. Diabetes Obes Metab 2018 Aug;20(8):2009-2013. [doi: 10.1111/dom.13307] [Medline: 29582538]
11. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials. Diabetes Care 2016 Nov;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]
12. Cui M, Wu X, Mao J, Wang X, Nie M. T2DM self-management via smartphone applications: a systematic review and meta-analysis. PLoS ONE 2016;11(11):e0166718. [doi: 10.1371/journal.pone.0166718] [Medline: 27861583]
13. Eze ND, Mateus C, Cravo Oliveira Hashiguchi T. Telemedicine in the OECD: an umbrella review of clinical and cost-effectiveness, patient experience and implementation. PLoS One 2020;15(8):e0237585. [doi: 10.1371/journal.pone.0237585] [Medline: 32790752]
14. Hazime D, Burner E. Social support via internet communication technology for diabetes self-management: a scoping review. Mhealth 2024;10:18. [doi: 10.21037/mhealth-23-34] [Medline: 38689617]
15. Shan R, Sarkar S, Martin SS. Digital health technology and mobile devices for the management of diabetes mellitus: state of the art. Diabetologia 2019 Jun;62(6):877-887. [doi: 10.1007/s00125-019-4864-7] [Medline: 30963188]
16. Burner E, Hazime D, Menchine M, et al. mHealth social support versus standard support for diabetes management in safety-net emergency department patients: randomized phase-III trial. JMIR Diabetes 2025 Apr 23;10:e56934. [doi: 10.2196/56934] [Medline: 40266665]
17. Burner E, Mercado J, Hernandez-Saenz A, et al. Design and patient characteristics of the randomized controlled trial TExT-MED + FANS a test of mHealth augmented social support added to a patient-focused text-messaging intervention for emergency department patients with poorly controlled diabetes. Contemp Clin Trials 2019 May;80:1-8. [doi: 10.1016/j.cct.2019.03.003] [Medline: 30878623]
18. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010;32(1):56-69. [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
19. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS One 2017;12(3):e0173160. [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
20. Bonoto BC, de Araújo VE, Godói IP, et al. Efficacy of mobile apps to the care of patients with diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. JMIR mHealth uHealth 2017 Mar 1;5(3):e4. [doi: 10.2196/mhealth.6309] [Medline: 28249834]




XSL•FO
RenderX

21. Odnoletkova I, Ramaekers D, Nobels F, Goderis G, Aertgeerts B, Annemans L. Delivering diabetes education through nurse-led telecoaching: cost-effectiveness analysis. PLoS One 2016;11(10):e0163997. [doi: 10.1371/journal.pone.0163997] [Medline: 27727281]
22. Yang F, Wang Y, Yang C, Hu H, Xiong Z. Mobile health applications in self-management of patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis of their efficacy. BMC Pulm Med 2018 Sep;18(1):147. [doi: 10.1186/s12890-018-0671-z] [Medline: 30180835]
23. Albanese AM, Huffman JC, Celano CM, et al. The role of spousal support for dietary adherence among type 2 diabetes patients: a narrative review. Soc Work Health Care 2019 Mar;58(3):304-323. [doi: 10.1080/00981389.2018.1563846] [Medline: 30596355]
24. Helgeson VS, Mascatelli K, Seltman H, Korytkowski M, Hausmann LRM. Implications of supportive and unsupportive behavior for couples with newly diagnosed diabetes. Health Psychol 2016 Oct;35(10):1047-1058. [doi: 10.1037/hea0000388] [Medline: 27280364]
25. Epple C, Wright AL, Joish VN, Bauer M. The role of active family nutritional support in Navajos' type 2 diabetes metabolic control. Diabetes Care 2003 Oct;26(10):2829-2834. [doi: 10.2337/diacare.26.10.2829] [Medline: 14514587]
26. Fisher EB, Boothroyd RI, Elstad EA, et al. Peer support of complex health behaviors in prevention and disease management with special reference to diabetes: systematic reviews. Clin Diabetes Endocrinol 2017;3(1):4. [doi: 10.1186/s40842-017-0042-3] [Medline: 28702258]
27. Owens GM. Gender differences in health care expenditures, resource utilization, and quality of care. J Manag Care Pharm 2008 Apr;14(3 Suppl):2-6. [doi: 10.18553/jmcp.2008.14.S6-A.2] [Medline: 18439060]
28. Vaidya V, Partha G, Karmakar M. Gender differences in utilization of preventive care services in the United States. J Womens Health (Larchmt) 2012 Feb;21(2):140-145. [doi: 10.1089/jwh.2011.2876] [Medline: 22081983]
29. Zeng Y, Wan Y, Yuan Z, Fang Y. Healthcare-seeking behavior among chinese older adults: patterns and predictive factors. Int J Environ Res Public Health 2021 Mar 14;18(6):2969. [doi: 10.3390/ijerph18062969] [Medline: 33799366]
30. Toschi E, Fisher L, Wolpert H, Love M, Dunn T, Hayter G. Evaluating a glucose-sensor-based tool to help clinicians and adults with type 1 diabetes improve self-management skills. J Diabetes Sci Technol 2018 Nov;12(6):1143-1151. [doi: 10.1177/1932296818791534] [Medline: 30060682]
31. Waki K, Fujita H, Uchimura Y, et al. DialBetics: a novel smartphone-based self-management support system for type 2 diabetes patients. J Diabetes Sci Technol 2014 Mar;8(2):209-215. [doi: 10.1177/1932296814526495] [Medline: 24876569]
32. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? JAMA 2013 Dec;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
33. Zhai YK, Zhu WJ, Cai YL, Sun DX, Zhao J. Clinical- and cost-effectiveness of telemedicine in type 2 diabetes mellitus: a systematic review and meta-analysis. Medicine (Baltimore) 2014 Dec;93(28):e312. [doi: 10.1097/MD.0000000000000312] [Medline: 25526482]
34. Rose AJ, Timbie JW, Setodji C, Friedberg MW, Malsberger R, Kahn KL. Primary care visit regularity and patient outcomes: an observational study. J Gen Intern Med 2019 Jan;34(1):82-89. [doi: 10.1007/s11606-018-4718-x] [Medline: 30367329]
35. Ho V, Metcalfe L, Dark C, et al. Comparing utilization and costs of care in freestanding emergency departments, hospital emergency departments, and urgent care centers. Ann Emerg Med 2017 Dec;70(6):846-857. [doi: 10.1016/j.annemergmed.2016.12.006] [Medline: 28262320]


## Abbreviations

**$A_{1c}$:** glycated hemoglobin
**COPD:** chronic obstructive pulmonary disease
**ED:** emergency department
**EMR:** electronic medical record
**FANS:** Family and Friends Network Support
**mHealth:** mobile health
**TExT-MED:** trial to examine text-messaging in emergency patients with diabetes

*Edited by N Cahill, T Cordoso; submitted 08.Nov.2024; peer-reviewed by F Al-Ani, YC Liu; revised version received 31.Oct.2025; accepted 31.Oct.2025; published 27.Feb.2026.*

*Please cite as:*
*Hazime D, Raffi L, Burner E*
*Impact of mHealth-Augmented Social Support on Health Care Use Among Patients With Diabetes: Secondary Analysis of the TExT-MED+FANS Trial*
*Interact J Med Res 2026;15:e65113*
*URL: https://www.i-jmr.org/2026/1/e65113*
*doi:10.2196/65113*



XSL•FO
RenderX

© Danielle Hazime, Liza Raffi, Elizabeth Burner. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 27.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Impact of a New Virtual Health Humanities Approach for Collaborative Online International Learning: Mixed Methods Study

Heather Meredith Gaunt[1], BMUS, BA, PhD; Kwang Meng Cham[2], PhD; Florian Knothe[3], MA, PhD; Chee Han Lim[4], PhD; Rosie Shea[5], MBBS; Hui Ting Chng[6], PhD

[1]Museums and Collections Department, Chancellery, University of Melbourne, Grattan Street, Parkville, Australia

[2]Department of Optometry and Vision Sciences, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, Australia

[3]University Museum and Art Gallery, School of Humanities, Faculty of Arts, University of Hong Kong, Hong Kong, China

[4]NUS College, National University of Singapore, Singapore, Singapore

[5]Department of Medical Education, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, Australia

[6]Department of Pharmacy and Pharmaceutical Sciences, Faculty of Science, National University of Singapore, Singapore, Singapore

**Corresponding Author:**
Heather Meredith Gaunt, BMUS, BA, PhD
Museums and Collections Department, Chancellery, University of Melbourne, Grattan Street, Parkville, Australia

## Abstract

**Background:** This study presents a novel virtual health-humanities approach to collaborative online international learning, developed and delivered between 2023 and 2025. The approach, which we describe for the purposes of this study as a "Health Humanities Global Classroom" (HHGC), aimed to raise health professions students' awareness of intercultural perspectives, potentially fostering more interculturally attuned health professionals. In the HHGC, students across 3 international universities participated synchronously in a virtual object-based learning seminar using shared teaching expertise and digitized artworks and museum objects drawn from the universities' diverse art and cultural collections.

**Objective:** This mixed methods research study aimed to assess the HHGC's impact on encouraging a deeper awareness of intercultural perspectives in participants.

**Methods:** A mixed methods study was conducted, involving 58 health professions students from universities in Australia, Singapore, and Hong Kong in February 2025 and March 2025. All students participated in a 2-hour online seminar with the theme of global health and well-being, during which they undertook guided exploration of virtual museum objects and artworks. Data on students' seminar experiences were collected through presurveys and postsurveys, written reflections created during the seminar, and online focus groups conducted immediately afterward.

**Results:** All 58 students participated in the focus groups. Among the participants, 84% (49/58) completed the postseminar survey; 96% (47/49) agreed that the seminar increased their awareness of how they communicate and exchange meanings with peers from diverse cultures; 98% (48/49) agreed that engaging with the museum objects in the seminar with peers heightened their awareness of diverse perspectives; and 86% (42/49) agreed that the seminar increased their awareness of local health in interprofessional contexts. Four primary themes were identified through qualitative analysis, including (1) key learning outcomes (intercultural, analytical, and empathic capacities, and knowledge acquired), (2) enabling factors (pedagogy, environment, and peer interaction), (3) barriers to learning (pedagogical approach, learning resources, time factors, and virtual constraints), and (4) novel and distinctive experiences (international, arts-integrated learning).

**Conclusions:** This study provides early evidence for an innovative and accessible pedagogical approach. The HHGC is promising as a valuable educational activity, with positive student engagement in a safe environment for intercultural dialogue mediated by virtual museum objects and artworks. This approach may be useful to enhance participants' awareness of cultural humility.

***(Interact J Med Res 2026;15:e84474)*** doi:10.2196/84474

**KEYWORDS**

internationalization; intercultural learning; intercultural perspectives; e-learning; collaborative online international learning; COIL; global classroom; health humanities; medical humanities; object-based learning; OBL; museum-based health education; MB-HE; cultural humility; cultural diversity; mixed methods



XSL•FO
RenderX

## Introduction

### Background

Providing culturally sensitive and appropriate care is a cornerstone of modern global health equity. Medical and educational institutions increasingly seek curricula that foster intercultural competence [1-6]. While there is currently no consensus on the definition of "intercultural competence," it is widely understood as the "ability to communicate effectively and appropriately in intercultural situations based on one's knowledge, skills, and attitudes" [7,8]. Additionally, as a pedagogical objective, the acquisition of "intercultural competence" betrays an assumption that culture is a finite body of knowledge to be mastered, a view that risks reducing complex human worlds to a static checklist of traits [7].

A combination of methods such as lectures with simulations, role-playing, and community engagement with diverse populations has been found to enhance intercultural competence [9]. Experiential learning, such as international placements, has also made a positive impact on the development of cultural competence among health professional students [10]. However, these incur financial costs and logistical barriers, making them inaccessible for many. The reach of such interventions is further limited by packed, discipline-specific health curricula that relegate humanistic approaches that cultivate essential skills like empathy, observation, and comfort with ambiguity to their periphery. This creates a critical gap: a need for scalable, accessible, and pedagogically innovative methods to mainstream intercultural competence learning.

To address these pedagogical challenges, we designed a new humanities-based methodology which could be used as an intervention in an existing health professions curriculum. Our approach, which we describe as a "Health Humanities Global Classroom" (HHGC), builds on the known potential of 3 pedagogical approaches, specifically museum-based health education (MB-HE), object-based learning (OBL), and collaborative online international learning (COIL).

### MB-HE

MB-HE uses carefully facilitated engagement with art and museum collections to support the development of transferable skills and ways of thinking among health professionals, such as visual observation, tolerance for ambiguity, awareness of bias and diversity of perspectives, empathy, and self-reflection. By connecting these experiences in the museum to clinical practice, the approach aims to promote more humanistic, patient-centered care, and simultaneously support well-being and reduce burnout [11,12]. MB-HE activities are commonplace in many universities' health education curricula around the world and are extensively reported as an effective pedagogical approach [11,12]. Particularly since the COVID-19 pandemic, this approach has been translated into online learning environments with varied success [13-15]. Online collaborative intercultural peer group activity in the global health educational context has been beneficial for student learning by complementing existing teaching in global health [16].

### OBL

OBL is a student-centered, experiential pedagogy that uses direct engagement with artworks, museum objects, and artifacts, and other types of material culture such as archives or scientific instruments, to stimulate observation, enquiry, and knowledge construction [17]. Differing from MB-HE, it is applied in any disciplinary context and, in its simplest form, is taught as a small-group "hands-on" investigation of a single object, where meaning or interpretation is built through discussion [17]. OBL has been adopted globally and is increasingly recognized as a powerful tool to promote ethical considerations of material culture, collections, and power [18].

### COIL

COIL is an internationalized educational approach in which educators from different countries and universities co-design and cofacilitate curriculum-integrated learning experiences that bring students together through real-time online interaction and collaborative inquiry. By engaging in sustained intercultural dialogue and joint academic tasks over the course of a "project" or part of a curriculum, students can potentially develop intercultural competence and understandings of global perspectives, as well as build collaborative and teamwork skills [19,20]. As Hackett et al [21] noted, measuring whether COIL programming successfully improves participants' intercultural competence is difficult, so further investigation is required to clarify its effectiveness in this area [21].

### Objectives

Building on the aforementioned pedagogies, we hypothesized that a teaching and learning approach that privileges awareness of multiple perspectives on a relevant health topic across a global context, using OBL with diverse cultural artifacts as mediating objects, in a virtual COIL context, would allow students to develop a deeper awareness of aspects of intercultural competence. We shaped our HHGC methodology to create a "third space" for intercultural dialogue [22], where learning is a situated activity and where knowledge is coconstructed within a temporary "community of practice" formed by the international student group [23]. We speculated that, through the OBL approach, the shared (virtual) artwork or cultural artifact would function as a neutral ground upon which students could negotiate complex meanings and explore sensitive topics without feeling personally exposed. The approach would thus foster a mode of inquiry that was analytical and creatively engaged. We undertook a research study that aimed to give us a deeper understanding of participants' experiences of our HHGC methodology, and the potential of the methodology to support students' development of intercultural competence.

## Methods

### Overview

The development phase of our HHGC methodology has been published previously in brief [24]. In this article, we report the results of a pilot delivery of the HHGC methodology conducted in early 2025 with health students from 3 international universities: the University of Melbourne (UoM), the University of Hong Kong (HKU), and the National University of Singapore

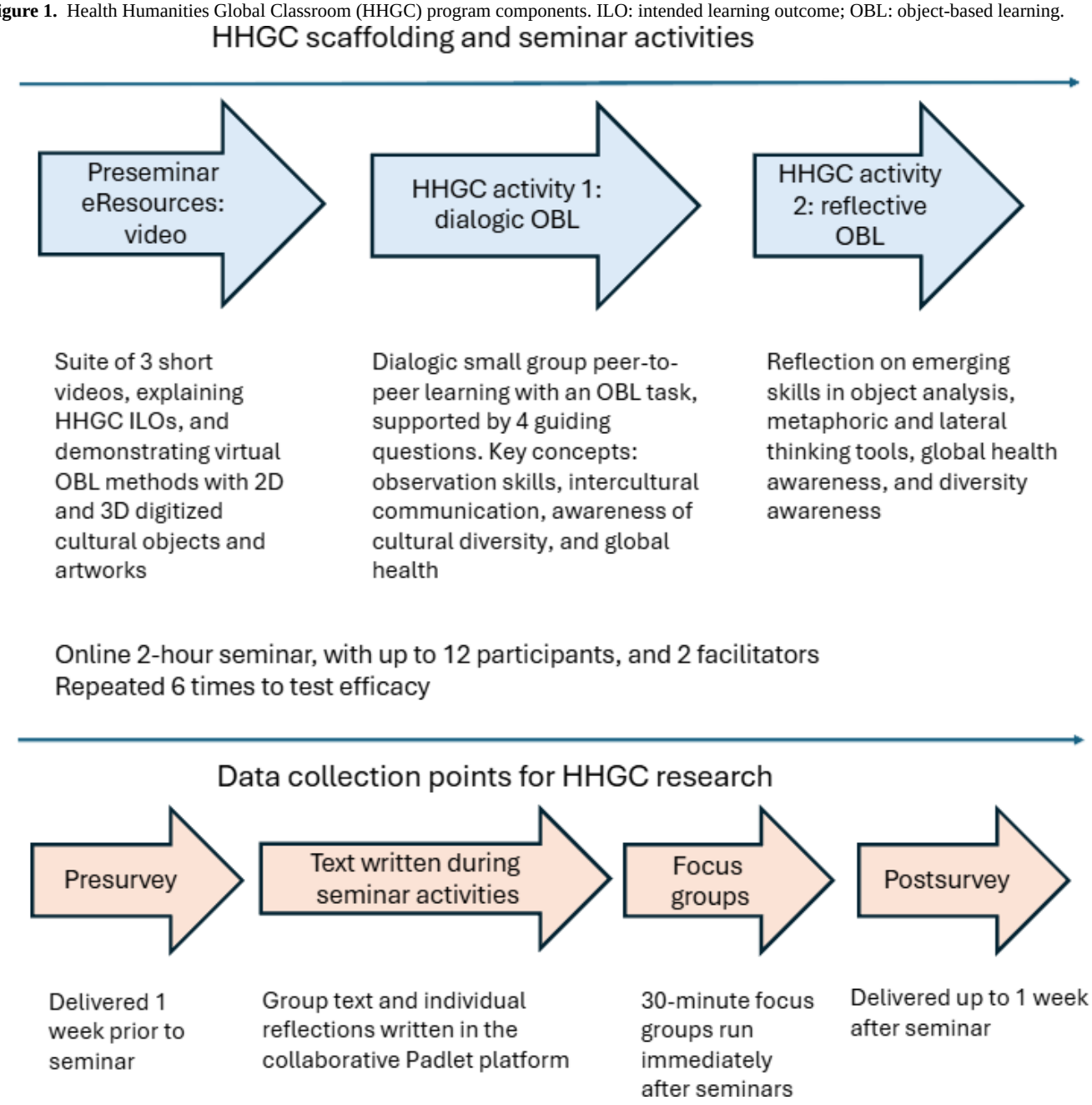


XSL•FO
RenderX

(NUS). The full HHGC project, including the learning components and associated research to understand its impact, is shown in Figure 1, and described in detail below.

The identical HHGC seminar was delivered in English and ran for 2 hours. The seminar was delivered on 6 occasions, each attended by between 8 and 11 participants, and each participant took part in a single session. To maximize the cultural diversity of the learning experience, we allocated students as evenly as possible from each university site, according to student availability. The mix of students in each seminar is shown below (Table 1).

**Figure 1.** Health Humanities Global Classroom (HHGC) program components. ILO: intended learning outcome; OBL: object-based learning.






XSL•FO
RenderX

**Table .** Distribution of participants across 6 Health Humanities Global Classroom (HHGC) seminars (by university).

| Seminar repeat delivery, n | Student from UoM[a], n | Student from HKU[b], n | Student from NUS[c], n | Total seminar participants, n |
|---|---|---|---|---|
| 1 | 4 | 1 | 5 | 10 |
| 2 | 4 | 1 | 3 | 8 |
| 3 | 3 | 1 | 5 | 9 |
| 4 | 3 | 2 | 6 | 11 |
| 5 | 2 | 1 | 6 | 9 |
| 6 | 5 | 2 | 4 | 11 |
| Total participants across all seminars: | | | | 58 |

[a]UoM: University of Melbourne.

[b]HKU: University of Hong Kong.

[c]NUS: National University of Singapore.

The seminars were facilitated by 2 members of the team, one with health professional expertise, and one with museum expertise, representing 2 different universities. Across the 6 occasions, all research team members contributed to facilitation at least once. Facilitators closely followed a script cocreated by all members of the team to ensure uniformity of the participants' experience. The seminar theme was broad, with discussions shaped around thoughtful comparisons of health and well-being in the contexts of each university's local and national environment. Using Anderson and Krathwohl's revision of Bloom's Taxonomy [25], our intended learning outcomes for the pilot HHGC (encompassing Levels 2 to 6 of the Taxonomy) were as follows: (1) use effective communication with students from different disciplinary backgrounds and cultures (Level 2: Understand and Level 3: Apply); (2) apply effective teamwork skills with students from different disciplinary backgrounds and cultures (Level 3: Apply and Level 4: Analyze); (3) develop an understanding of different cultural and global perspectives on health care (Level 2: Understand and Level 4: Analyze); and (4) build observational, creative, and critical thinking skills to discuss the presented artwork or museum object (Level 4: Analyze, Level 5: Evaluate, and Level 6: Create).

The OBL content for discussion consisted of a virtual group (e-Gallery) of 24 artworks and museum objects, which we selected in equal numbers from the cultural collections of the 3 universities (Figure 2).

Artworks were selected by the museum professionals on the team, who worked closely together online to curate the final group to provide a balance of cultural and historical contexts of the artworks' production and a rich potential to trigger discussions about the human body, health, or well-being.

Facilitators formed breakout-room teams of 3 to 4 participants at the start of each seminar, aiming, where possible, for each team to include at least 1 student from each university site. In the facilitated OBL activities, teams selected 1 object from the e-Gallery for in-depth exploration and discussion. Four guiding questions scaffolded the team discussions, encouraging students to move logically through a process: to undertake close observation of the object; to build a detailed, collective description; to consider and share their emotional responses; and to reflect on how their cultural perspectives might shape their observations. Teams documented their close observations on the collaborative online platform Padlet, including how they connected them to the HHGC's overarching theme of health [26]. In the final part of the seminar, students each selected a new object and crafted short reflective statements that captured their understanding of their object, and how the group discussion may have influenced or changed their analysis. Participants were encouraged throughout the seminar to use an enquiring approach, to make direct or lateral connections between the artworks and the broad theme of global health, and to contribute these actively to the group discussion.



XSL•FO
RenderX

**Figure 2.** Screen capture of a section of the e-Gallery resource of cultural objects and artworks.

## Ethical Considerations

We designed a mixed methods research study to evaluate the impact of the HHGC on intercultural learning for participants. Ethical approval has been granted by all 3 universities where students were participating in this project: The University of Melbourne Office of Research Ethics and Integrity (reference: 2024-28502-54028-3; approval date: May 31, 2024), National University of Singapore Institutional Review Board (reference: NUS-IRB-2024-467; approval date: September 27, 2024), and The University of Hong Kong Research Ethics Institutional Review Board (protocol: UW 24-583; approval date: October 25, 2024). All participants provided informed consent through Qualtrics, and participation was entirely voluntary. All surveys were anonymous, and participants were not identifiable. All data from focus group recordings and text data created by participants during the HHGC activities were anonymized and deidentified to safeguard participant information. Participants were reimbursed for their time with a US $65 gift voucher.



XSL•FO
RenderX

## Participants

Participants in the HHGC were diverse, drawn from students studying at different levels and in different curriculum contexts at the 3 universities. A convenient sampling approach was undertaken to recruit participants at each of the university sites, via a recruitment email posted on the learning management system from members of the research team to specific cohorts (for example, students studying an elective "Medical Humanities" topic at UoM). In the presurvey, delivered to participants via Qualtrics up to 2 weeks prior to the scheduled seminar, we sought to understand the participants' demographic and contextual (social, cultural, and historical) details, to learn more about their previous experiences in online learning and museum-based learning, and their current perception of confidence relating to intercultural communication ability (Table 2).

**Table .** Presurvey questions, Health Humanities Global Classroom.

| Question number | Question | Response selection |
|---|---|---|
| 1 | Which university are you currently studying at? | University of Melbourne, University of Hong Kong, and National University of Singapore |
| 2 | Which of the following area(s) of health is the focus of your current tertiary study? Select one or more answers. | Medicine, nursing, pharmacy, optometry, dentistry, audiology, social work, and other |
| 3 | At what year level of your degree are you currently studying? | Years 1, 2, 3, 4, 5, and 6 |
| 4 | Have you participated in an online international learning experience previously? | Yes/no |
| 5 | If you answered yes to the question above, briefly describe that experience. | Free-text answer |
| 6 | Have you participated in a museum-based learning experience as part of your tertiary education? | Yes/no |
| 7 | If you answered yes to the question above, briefly describe that experience. | Free-text answer |
| 8 | What language/s do you speak at home? | Free-text answer |
| 9 | Rate your response to the following question: "I am confident in my ability to communicate and exchange meanings with colleagues from diverse cultures in a variety of contexts." | Extremely confident, very confident, somewhat confident, not very confident, and not at all confident |

Focus groups of approximately 30 minutes were conducted and recorded on Zoom following each seminar. We chose focus groups so that participants could build on each other's ideas in a collaborative context, continuing the open and safe environment for sharing that had already been created in the seminar [27]. Focus groups were facilitated by members of the research team who had not taught in the preceding seminar; each facilitator followed a script with ethics-approved questions and prompts (Textbox 1). Each focus group had 3 to 6 participants, depending on the total number attending the seminar; the focus groups ran simultaneously in new, separate Zoom meetings beginning ten minutes after the seminar concluded. Focus-group participants were predetermined during the allocation of seminar attendees, with the aim of including as even a distribution from each university as possible. During the focus groups, participants were asked questions designed to explore how they individually experienced the seminar's contextual factors. Focus groups' transcriptions were anonymized, and each participant was given a code number.

**Textbox 1.** The presurvey questions delivered to students who indicated interest in participating in the health humanities global classroom seminar.

- What was your experience of participating in this seminar with international peers?
- What was the experience of exploring the global health theme in this specific seminar context like for you?
- What was your experience of sharing your personal perspectives and observations during the seminar?
- What was your experience of hearing the perspectives of the other students during the seminar?
- What was the experience of working with virtual museum objects like for you?
- Has this seminar been a unique experience for you in any way?

## Analysis

As noted in Figure 1, we collected quantitative data at the beginning and end of the data-collection phase using surveys, and qualitative data midstudy through in-seminar reflective texts on Padlet, and postseminar focus groups. Our surveys contained a mix of qualitative and quantitative items. We analyzed quantitative survey data in Qualtrics, reporting



XSL•FO
RenderX

percentages for categorical variables and means for Likert-scale items. GraphPad Prism 6.07 (GraphPad Software Inc) was used for graphical presentation of the data. We reported only descriptive statistics (percentages and means) because this pilot study was an exploratory evaluation of a novel pedagogical intervention, with a modest convenience sample and immediate postsession measurement. In this paper, we did not include data or analysis from the in-seminar reflective texts; that material will be included in a future paper that focuses on the impact of our virtual OBL approach, addressing the technical challenges and opportunities of the pedagogy. Qualitative data from the focus groups were analyzed thematically to categorize the major themes, using the qualitative data analysis software NVivo (version 15; Lumivero). We used a qualitative data analysis approach described by Fereday and Muir-Cochrane for text-based data [28]. An initial codebook was developed by HMG. Coding was legitimated when RS, HTC, and KMC (representing 2 of the 3 participating sites) each independently checked the coding on 4 focus-group transcripts. The platform Spark AI (UoM) was used by HMG to run queries across the qualitative data set of focus groups to explore potential additional subcodes. Themes and subthemes derived from the codes were discussed by the team via email and Zoom meetings. Any disagreements were resolved by HMG. The final consensus was shared with the full team as a report produced in NVivo, listing themes, related codes, theme descriptions, and statistics on the number of files and references for each.

Finally, we adopted a convergent mixed methods approach in which descriptive quantitative measures of participants' perceptions are interpreted alongside thematic qualitative data to provide complementary perspectives. Quantitative percentages describe the prevalence of positive perceptions across the cohort, while qualitative data explain how and why participants experienced these outcomes. In reporting, we therefore emphasized convergence (where high percentages aligned with recurrent qualitative themes) and complementarity (where qualitative data provided more nuance to the descriptive results).

All members of the research team were conscious of and reflected on their potential biases, recognizing that their diverse backgrounds as doctors, pharmacists, anthropologists, optometrists, clinical educators, museum curators, and educators may influence the interpretation of the results to some degree. Reflexive notes throughout the development and delivery of the pilot methodology and multiple discussions across the data analysis period ensured that coding decisions were anchored in the transcript data rather than researchers' *a priori* expectations, as much as possible.

## Results

### Participant Demographics

A total of 67 students were recruited for this study and completed the preseminar survey in the week prior to the seminars. Of the 67 students, 58 attended an HHGC seminar, and the 9 who did not attend were either unable to find a suitable seminar time or faced unforeseen personal circumstances. Because of survey anonymity, we were unable to identify which presurvey respondents did not continue in the study. All 58 students who attended an HHGC seminar participated in the focus groups; 84% (49/58) of seminar attendees completed the postseminar survey. The 67 participants in the presurvey ranged across first to fourth years of their current degree across the 3 university sites, with a few having recently graduated (Tables 3 and 4).

**Table .** Presurvey responses from 67 participants indicating location and current year of studies.

| University | First year; n | Second year; n | Third year; n | Fourth year; n | Recently graduated; n | Total responses; n |
|---|---|---|---|---|---|---|
| NUS[a] | 3 | 17 | 7 | 6 | 0 | 33 |
| HKU[b] | 3 | 3 | 1 | 3 | 0 | 10 |
| UoM[c] | 0 | 1 | 15 | 4 | 4 | 24 |
| Total | 6 | 21 | 23 | 13 | 4 | 67 |

[a]NUS: National University of Singapore.

[b]HKU: University of Hong Kong.

[c]UoM: University of Melbourne.



XSL•FO
RenderX

**Table .** Presurvey responses from 67 participants showing university site, and their current disciplinary focus/foci.

| Site | Medicine; n | Pharmacy; n | Nursing; n | Population and global health; n | Nursing, population and global health, and health and well-being; n | Pharmacy and health and well-being; n | Other, n | Total, n |
|---|---|---|---|---|---|---|---|---|
| NUS[a] | 4 | 21 | 3 | 0 | 0 | 3 | 2 | 33 |
| HKU[b] | 3 | 0 | 1 | 2 | 2 | 0 | 2 | 10 |
| UoM[c] | 23 | 0 | 0 | 0 | 0 | 0 | 1 | 24 |
| Total | 30 | 21 | 4 | 2 | 2 | 3 | 5 | 67 |

[a]NUS: National University of Singapore.

[b]HKU: University of Hong Kong.

[c]UoM: University of Melbourne.

All participants were fluent in English, and a large proportion were multilingual, speaking other languages at home, including Sinhalese, Mandarin, French, and Italian. From the presurvey data, 18% (12/67) had previously participated in an online international learning experience; 45% (30/67) had participated in a museum-based learning experience as part of their tertiary education, including single sessions in local history, art, or anatomy museums, and in some cases, more extensive museum experiences through enrollment in medical humanities courses.

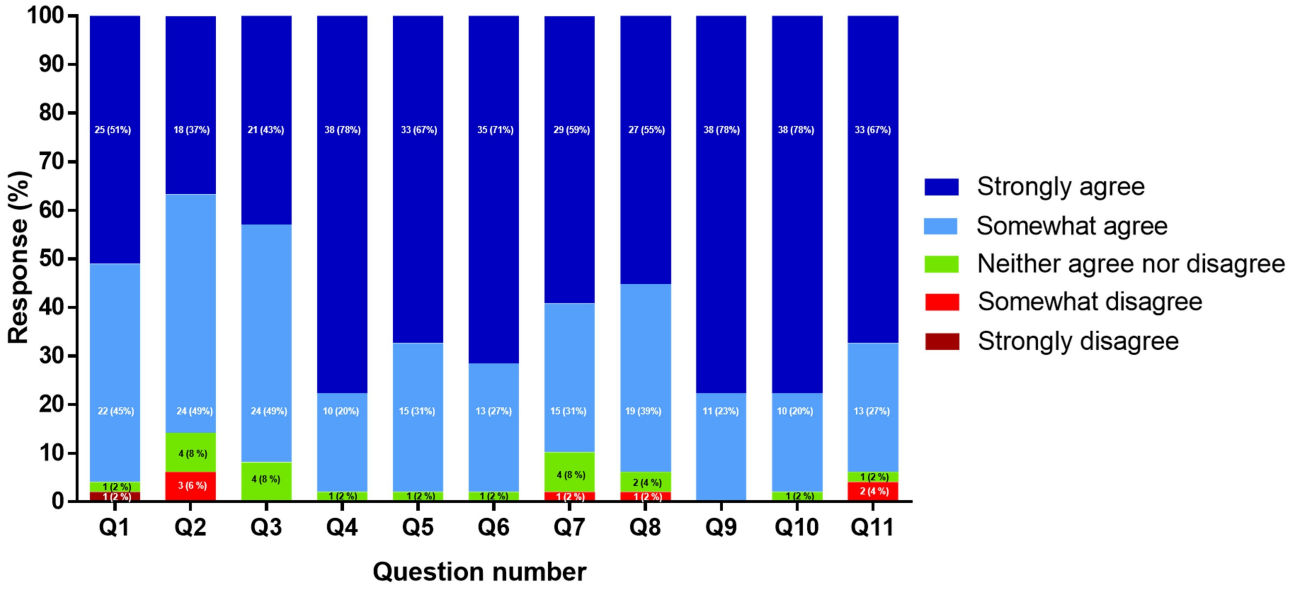
## Quantitative Analysis

From the 58 students, 84% (n=49) of seminar participants completed the postseminar survey (Figure 3). In this survey, 96% (47/49) of respondents agreed that the seminar had increased their awareness of how they communicated and exchanged meanings with peers from diverse cultures. Relating to health knowledge, 86% (42/49) agreed that the seminar had increased their awareness of local health in interprofessional contexts, while 92% (45/49) agreed that they had an increased understanding of health from a global perspective. Relating to the virtual OBL experience, 98% (48/49) agreed that engaging with the diverse museum objects was thought-provoking and that engaging with these objects in a peer learning environment heightened their awareness of diverse perspectives. The same proportion of participants also agreed that this COIL experience with international peers was an effective way to increase their awareness of intercultural competence.



XSL•FO
RenderX

**Figure 3.** Participant responses to postsurvey questions.



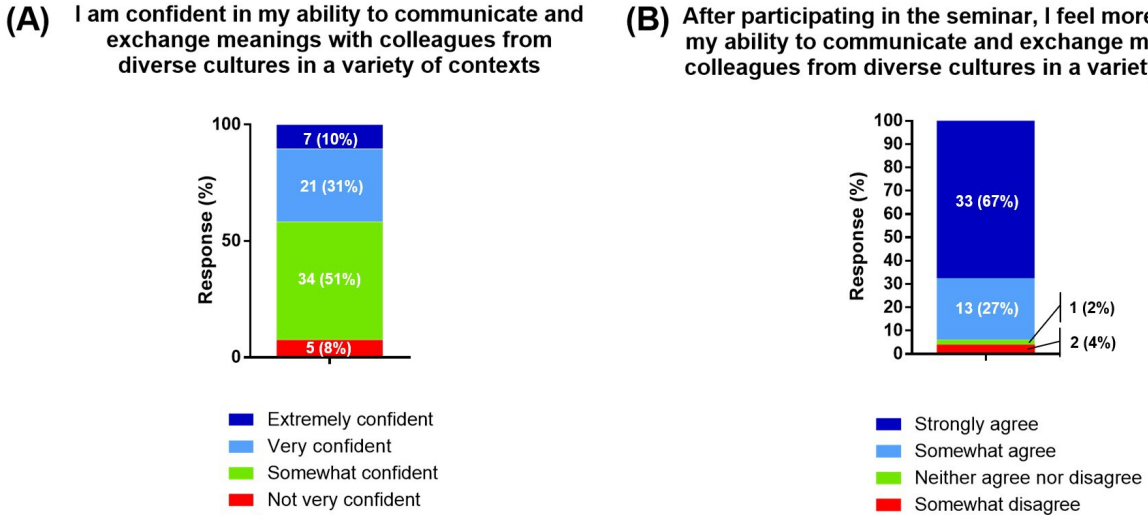

Q1 This seminar has increased my awareness of how I communicate and exchange meanings with peers from diverse cultures.
Q2 This seminar has increased my awareness of local health in interprofessional contexts.
Q3 This seminar increased my understanding of health from a global perspective.
Q4 Engaging with virtual museum objects from different universities was thought provoking.
Q5 Engaging with virtual museum objects in the seminar allowed me to explore the health theme with peers in a way that was novel to me.
Q6 Engaging with virtual museum objects in the seminar with peers heightened my awareness of diverse perspectives.
Q7 The online platform for interaction with the virtual museum objects was easy to use.
Q8 I found the pre-seminar online resources useful to prepare me for my seminar experience.
Q9 I found the Seminar Facilitators helped create an environment conducive to learning.
Q10 Participating in an online seminar with international peers is an effective way of increasing my awareness of intercultural competence.
Q11 After participating in the seminar, I feel more confident in my ability to communicate and exchange meanings with colleagues from diverse cultures in a variety of contexts.

In the presurvey and postsurvey, participants were asked to describe their immediate perception of confidence in communicating across cultures (Q9: presurvey; Q11: postsurvey). While our presurvey question asked about their current self-perception of confidence in communicating in intercultural contexts, the postsurvey question sought to identify any change in their perception of their confidence after seminar participation. Prior to the seminar, 92% of students who completed the survey were confident to some degree ("somewhat confident," "very confident," or "extremely confident"), with around half of that group sitting in the "somewhat confident" category. In the postsurvey, we saw that 94% (46/49) perceived an increase in their confidence through the experience (Figure 4).



XSL•FO
RenderX

**Figure 4.** Side-by-side graphic comparisons of (A) preseminar and (B) postseminar survey questions that explored participants' immediate self-perception of confidence in their intercultural communication ability.




## Thematic Analysis

Four key themes, with multiple subthemes, emerged from our thematic analysis of the textual data derived from transcriptions of the 12 focus groups and text-based responses in the postsurvey. The themes presented students' perceptions of their experiences of the seminar across three distinct areas, namely (1) their key learning outcomes, the enabling factors which allowed them to engage and learn effectively; (2) the barriers to learning they encountered; and (3) their articulation of what was unique or novel about the seminar experience. These themes are summarized in Table 5. The themes were listed in the order of frequency with which they occurred in the data. We also identified a fourth theme, which brought together their suggestions for improvements to the seminar, or future use of the methodology in health education. We did not include this fourth theme in the primary findings, as we considered that the data should be considered separately from that which reported their experience in the HHGC seminar itself.



XSL•FO
RenderX

**Table .** Students' experience of the Health Humanities Global Classroom (HHGC): qualitative themes and summary of subthemes.

| Key themes | Subthemes |
|---|---|
| 1. Key learning outcomes: intercultural, analytical, and empathic capacities, and knowledge acquired | • Self-perception of intercultural capacities enhanced, including new understandings of the influence of cultural lenses, effectiveness of diverse perspectives, and specific capacities regarding communication across cultures.<br>• Self-perception of skills and capacities enhanced, including observation skills, analytic style, increased empathy awareness, and importance of emotions.<br>• Knowledge acquired in health and global health, and personalized care.<br>• Knowledge acquired about art or culture. |
| 2. Enabling factors: pedagogy, environment, and peer interaction | • Exposure to diverse perspectives including from international peers.<br>• Engaging e-resources, and application of OBL[a] pedagogy.<br>• Safe, positive, comfortable, and inclusive learning environment.<br>• Collaborative peer discussions and exposure to peer expertise.<br>• Sense of commonality among peers.<br>• Using art to understand health<br>• Contribution of expert and supportive facilitators<br>• Encouragement of exploratory lateral thinking<br>• Small group size |
| 3. Barriers to learning: pedagogical approach, learning resources, time factors, and virtual constraints | • Limitations of e-resources as compared to experiencing physical artworks or objects.<br>• Unfamiliarity with the approach of using art to explore health.<br>• Challenge of virtual learning environment (as compared to a physical environment).<br>• Time constraints of a single seminar |
| 4. Novel and distinctive experiences: international, art-integrated learning | • Confluence of multiple HHGC factors (including virtual OBL with art, international peer exposure, and linked health theme, and so on).<br>• Difference of humanities learning environment compared with standard health education.<br>• Intersection of humanities and health<br>• Discussing art with international peers |

[a]OBL: object-based learning.

### Theme 1: Key Learning Outcomes

In terms of learning outcomes, participants expressed the acquisition of analytical, intercultural, and empathic capacities, and specific knowledge. The capacity most frequently identified by participants was an enhancement of their intercultural awareness. The second most common capacities we grouped as personal skills, such as observational skills, and critical or analytic thinking.

### Theme 2: Enabling Factors

The second theme helped us to understand the enabling factors that contributed to participants' learning. The highest-cited factor we identified was the students' exposure to diverse perspectives (including those from international peers). Engagement with e-resources and the OBL pedagogy was the second-highest factor. A positive and inclusive learning environment was the third most common factor identified.

### Theme 3: Barriers to Learning

This theme allowed us to understand the barriers students reported as interfering with their learning or engagement. The highest-cited factor focused on various limitations of the e-resources, mostly in comparison to a face-to-face experience with physical artworks or objects. The second highest barrier was unfamiliarity with the lateral approach of using art to explore health. However, all responses that cited this barrier included descriptions of how they were able to improve their confidence with the lateral approach during the seminar: what was initially experienced as a difficulty became a positive through the influence of other factors, including more time to work on a skill, learning from their peers, or effective facilitator interactions. As 1 participant observed:

> *For me, it was quite intimidating at first, because I didn't really have much experience with connecting humanities to health, and so I had to sort of come up with responses and opinions in such a short frame of time. … But once the flow, once the whole thing started rolling, and we each got more involved and then, it became smoother, the process of thinking and analyzing.* [2.1P2]

It is important to note that what some participants perceived as barriers were not perceived as barriers by others. For example, the online (Zoom platform) learning experience was identified by some participants as a highly effective way of engaging synchronously with international peers, while others found it to be less optimal than face-to-face learning.



XSL•FO
RenderX

### Theme 4: Novel and Distinctive Experiences

The fourth theme focused on the uniqueness of the HHGC experience. Participants identified a range of factors that were novel to them, such as the opportunity to explore art in relation to health topics, accessing virtual galleries of artworks scanned in three dimensions, and synchronous learning with international peers from diverse backgrounds. The most highly cited aspect was the confluence of more than three HHGC factors. One participant described this as “the whole package” (6.1P4). We compare the quantitative and qualitative findings in Table 6.



XSL•FO
RenderX

**Table .** Side-by-side convergence of Health Humanities Global Classroom (HHGC) quantitative and qualitative data.

| Analytic domain (theme) | Quantitative findings | Qualitative evidence (theme summary and representative quotes) |
|---|---|---|
| Key learning outcomes: intercultural, analytical, and empathic capacities, and knowledge acquired. | • 98% (48/49) agreed engaging with objects with peers heightened awareness of diverse perspectives.<br>• 98% (48/49) agreed the COIL[a] experience was effective to increase awareness of intercultural competence.<br>• 92% (45/49) increased understanding of health from a global perspective.<br>• 86% (42/49) increased awareness of local health in interprofessional contexts. | • "[I] understood much more clearly - through personal experience obtained from completing the exercises - how culture informs how we see the world and how layered and complicated that can be." (PS.P7)<br>• "I feel like …we … can also understand different perspectives from our own very different cultural contexts, like, from our international peers. And that I feel like such new perspective will allow us to gain a very holistic overview, which actually is something I feel like is very needed… given our current global landscape." (2.1P4)<br>• "I learned to link my thought processes on health-related concepts to the culture in which the item in question is produced, beyond simply making links to the item itself." (PS.P22) |
| Enabling factors: pedagogy, environment, and peer interaction | • 100% (49/49) agreed that the seminar facilitators helped create an environment conducive to learning.<br>• 98% (48/49) agreed engaging with virtual museum objects was thought-provoking.<br>• 90% (44/49) agreed that the online platform for interaction with the virtual museum objects was easy to use.<br>• 94% (46/49) agreed that the preseminar online resources were useful preparation for the seminar experience. | • "Engaging with peers from different linguistic and cultural backgrounds also strengthened cross-cultural communication skills and underlined the need for cultural sensitivity in global health work in the future." (PS.P38)<br>• "I was quite fascinated with the way that we did the seminar... the 3D image that has been created, give us a lot of details, so we can have a deeper discussion on the artwork... this is really important because it facilitated our discussion later." (3.2P4)<br>• "I definitely like the technology to support the exhibition, actually, because I think it's quite easy and convenient for us to access different resources and also we did like four or five questions, we have to post our ideas or comments on the Padlet, and that actually helps to connect with people from different cultural backgrounds or different locations." (6.2P4)<br>• "I felt there was a safe space to really share more about our personal perspectives about the artworks, and how they relate to health as a whole" (3.1P3). |
| Barriers to learning: pedagogical approach, learning resources, time factors, and virtual constraints | • 2% (1/49) somewhat disagreed that the online platform for interaction with the virtual museum objects was easy to use.<br>• 2% (1/49) somewhat disagreed that the preseminar online resources were useful preparation for the seminar experience. | • "I think scale plays a really big role in how you emotionally respond to something, and that is something that's lost through the digital gallery, unfortunately." (1.1P2)<br>• "[It was] certainly difficult sometimes to [sic] looking at some of the pieces and trying to think of how it relates to health, and yeah, just bringing it back to the theme." (6.2P1) |



XSL•FO
RenderX

| Analytic domain (theme) | Quantitative findings | Qualitative evidence (theme summary and representative quotes) |
|---|---|---|
| Novel and distinctive experiences: international, art-integrated learning | • 98% (48/49) agreed that engaging with virtual museum objects in the seminar allowed me to explore the health theme with peers in a way that was novel to them. | • "But I think, generally, the virtual experience is really amazing, that you can actually get it all that work from different like different authors, different cultural backgrounds, that allow us to actually integrate it into the health care or into the humanity concept. So I think, yeah, it's generally very, very unforgettable." (6.1P4)<br>• "Creativity is such a, like, nourishing thing for the human spirit… and it's something that in health care, you know, we don't spend much time reflecting on it, because we're so busy just getting the job done. So it's wonderful to have these experiences where we share our thoughts about the world, and look at it from a different perspective, like using art as the, sort of, talking point, around which we can focus the discussion." (2.1P3) |

[a]COIL: collaborative online international learning.

## Discussion

### Principal Findings

A novel virtual health humanities approach for collaborative and international learning among health professional students, the HHGC, was developed. While the HHGC draws on established pedagogies (MB-HE, OBL, and COIL), its contribution to educational practice lies in their structured synthesis as a synchronous, virtual, cross-institutional seminar that uses shared, curated cultural objects as mediating tools. This integrated format creates an accessible "third-space" for intercultural dialogue that leverages material culture as a neutral focal point, supports small-group inquiry across institutions, and is designed for scalability and sustainable access. In this way, the HHGC offers a practicable model for embedding humanities-informed, intercultural learning into curricula where face-to-face exchanges or international placements are infeasible.

We found that the HHGC approach raised participants' awareness of diverse cultural perspectives and provided stimulating opportunities to consider intercultural contexts through a humanities lens. We believe the perceived benefits of the seminar can be attributed to its unique design, which addresses gaps in traditional health education through a form of situated learning. By creating a supported and safe environment for dialogue mediated by virtual museum objects and artworks, the HHGC provided a psychologically safe environment where knowledge could be coconstructed, rather than merely transmitted [29]. As we had hypothesized, the HHGC structure allowed students to explore sensitive topics without the defensiveness that can arise in unmediated conversations, mitigating the risk of stereotyping. The objects served as a neutral focal point, enabling students to discover both surprising commonalities and profound cultural differences in a low-stakes context. As one student noted, this allowed for a more holistic overview. Several participants closely connected the HHGC's discussions and focused analysis of artworks to their professional clinical experiences and appreciated the power of developing one's ability to observe closely as they recognized these as transferable skills. This personal experience was coupled with the newly gained knowledge for comparing observations with group members of different cultural backgrounds and appreciating these differences, shedding bias, and establishing a more neutral yet holistic assessment. Some reflections written by participants offered suggestions for curriculum integration in broader health education contexts and outlined some potential applications in health care delivery. This will be valuable as we develop the methodology further.

We were particularly struck by how intellectually and creatively engaged the students typically were during the seminar, despite their perceived barriers such as unfamiliarity with OBL or the virtual learning context. As facilitators, we noticed how breakout room conversations were consistently active and productive, and how participants were confident to express new ideas about artwork interpretations, often drawing on their emotional responses to the artists' themes or the appearance of the artworks. In the data, we found many instances where participants used adverbial intensifiers in their descriptions of the value or personal impact of the experience, such as "pretty eye opening," "very surprising," and "genuinely, really, shocked" (in a positive sense). We consider that this suggests that participants experienced moments of reflexive self-confrontation at points in the seminar, potentially signaling the beginning of greater awareness and humility that is foundational to authentic intercultural engagement [30].

Considering our experiences in delivering the HHGC, and the data from our study showing the students' experiences, we posit that the methodology could be valuable in promoting cultural humility in students. Cultural humility, as originally described by Tervalon and Murray-García [30], is a lifelong process of self-evaluation and critique of one's own cultural biases and limitations, oriented toward redressing power imbalances in the clinical encounter [31]. Bora et al [32] include cultural humility



XSL•FO
RenderX

as one of the key elements when setting up cross-cultural dialogue in psychiatry training, along with open-ended inquiry, reflective practice, active listening, didactic education, and a collaborative approach. As a distinct element in a suite of approaches such as those described by Bora et al [32], the HHGC could serve as a stimulating humanities-based context to access diverse cultural perspectives. Our HHGC methodology centers ambiguity, interpretation, and highlights the validity of multiple perspectives. For these reasons, it may be a fruitful pedagogical tool in a larger approach that pushes back against the "hidden curriculum" of biomedicine, which often privileges certainty and a single, objective truth [31,33].

### Strengths and Limitations

We believe this study has provided evidence for the potential of our novel pedagogical approach to support participants' development toward the important educational goal of increased intercultural competence and awareness. The curriculum, through being delivered in virtual learning environments, supports the goals of sustainable and equitable access to internationalizing experiences for learners in the health professions. However, our study of this pedagogy faced limitations. Importantly, the outcomes reported in this study reflect participants' immediate, self-reported perceptions of learning following a single 2-hour seminar, rather than direct measures of demonstrated competence. These findings, therefore, indicate perceived changes in awareness, confidence, and analytic orientation, rather than objective or sustained changes in intercultural skills. Future work should assess demonstrated competence using validated performance-based measures and longitudinal follow-up to determine whether perceived gains translate into durable, observable changes in behavior and clinical practice. Further limitations include self-selection bias from the recruited cohort: the sample reflects students who were available and willing to participate in a humanities-oriented virtual seminar and may therefore be more favorably disposed to art-integrated learning and online learning than the broader student population. The geographic representation, including participants drawn from universities in Australia, Singapore, and Hong Kong, while offering some diversity, does not capture the full range of global health education contexts and cultures, while the recruitment of primarily English-fluent participants limits generalizability. These sampling constraints should be borne in mind when considering the transferability of results to other settings.

Finally, the application of the HHGC methodology into health education contexts has specific development and delivery factors, requiring expert facilitators who have the skills and inclination to collaborate across the disciplines of health professional education and museum-based tertiary education, as well as access to compelling digital art resources. Additional challenges for delivery into the curriculum in health education include complex scheduling for cohorts learning together across multiple international sites and time-zones, a characteristic common to COIL educational programming, but it is acutely the case in the very full curriculum common in health education contexts.

### Future Directions

We believe that the methodology of the HHGC has the potential to be translated into a range of health education contexts, by shaping the OBL context around a range of health themes, or broader topics such as sustainability in health care provision, or interprofessional collaboration and leadership. The HHGC methodology may be embedded into educational activities in curricula, either as short single-seminar interventions or as multiweek COIL projects. However, further research on the impact of these embedded experiences, on their longitudinal impact on students' cultural awareness or development of cultural humility, and on how they might shape students' interactions with their patients in clinical environments would be highly valuable to identify areas in which the methodology would be most effectively implemented and where it can be improved. Implementation research could also compare virtual versus in-person OBL experiences to clarify the trade-off between accessibility and embodied engagement with objects.

### Conclusions

The proposed HHGC methodology has the potential to be implemented in health professions learning, where there is an educational need to promote intercultural capabilities. This pilot study suggests that the HHGC could offer a sustainable, scalable, and replicable approach for educators as a resource to educate future health care professionals with the critical personal and intercultural awareness necessary to provide humane care in a globalized world.

## Acknowledgments

The authors acknowledge the central contribution to the development and delivery of the entire Health Humanities Global Classroom project, of Prof Kit Wa Chan, Yueh Siang Chang, Eugene Koh, and Ben Loveridge. The authors used computer tools to identify and organize relevant scholarly work. The University of Melbourne Spark AI platform was used in this study as a check on aspects of the first round qualitative coding by HMG, prior to loading codes into NVivo (Lumivero) for subsequent coding. Specifically, AI was asked to find all quotes it considered to match the first-round codes across the whole dataset. These quotes were then compared with the original quotes found by HMG. AI was asked to suggest additional subcodes for the dataset, and these subcodes were considered by HMG as a comparison to subcodes created by HMG and other members of the research team.

## Funding

This project received seed funding from Universitas 21 (no grant number), and The University of Melbourne (Chancellery International Global Classrooms Seed Grants 2023, no grant number).



XSL•FO
RenderX

## Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

## Authors' Contributions

All authors made substantial contributions to the design of the study and the acquisition of data. SMG, HTC, KMC, and RS made substantial contributions to the data analysis. All authors made substantial contributions to the interpretation of the data, drafting the work, and revising the work. All authors provided final approval of the version to be published. All authors agreed to be accountable for all aspects of the work, ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## Conflicts of Interest

None declared.

## References


1. Li S, Miles K, George RE, Ertubey C, Pype P, Liu J. A critical review of cultural competence frameworks and models in medical and health professional education: a meta-ethnographic synthesis: BEME Guide No. 79. Med Teach 2023 Oct;45(10):1085-1107. [doi: 10.1080/0142159X.2023.2174419] [Medline: 36755385]
2. Horvat L, Horey D, Romios P, Kis-Rigo J. Cultural competence education for health professionals. Cochrane Database Syst Rev 2014 May 5;2014(5):CD009405. [doi: 10.1002/14651858.CD009405.pub2] [Medline: 24793445]
3. Shaya FT, Gbarayor CM. The case for cultural competence in health professions education. Am J Pharm Educ 2006 Dec 15;70(6):124. [doi: 10.5688/aj7006124] [Medline: 17332850]
4. Constantinou CS, Andreou P, Nikitara M, Papageorgiou A. Cultural competence in healthcare and healthcare education. Societies 2022;12(6):178. [doi: 10.3390/soc12060178]
5. Klenner M, Mariño R, Pineda P, Espinoza G, Zaror C. Cultural competence in the nursing, dentistry, and medicine professional curricula: a qualitative review. BMC Med Educ 2022 Sep 20;22(1):686. [doi: 10.1186/s12909-022-03743-7] [Medline: 36127655]
6. Taylor G, Papadopoulos I, Dudau V, Maerten M, Peltegova A, Ziegler M. Intercultural education of nurses and health professionals in Europe (IENE). Int Nurs Rev 2011 Jun;58(2):188-195. [doi: 10.1111/j.1466-7657.2011.00892.x] [Medline: 21554292]
7. Deardorff DK. Identification and assessment of intercultural competence as a student outcome of internationalization. J Stud Int Educ 2006 Sep;10(3):241-266. [doi: 10.1177/1028315306287002]
8. Al-Haddad M, Lu PY. It takes two to tango: the “inter” in intercultural competence. Med Educ 2024 Jul;58(7):766-768. [doi: 10.1111/medu.15319] [Medline: 38290785]
9. Nkhoma G, Lim CX, Kennedy GA, Stupans I. Teaching and learning interventions designed to improve cultural competence in health profession students: a systematic review. Internet J Allied Health Sci Pract 2023;21(3):2. [doi: 10.46743/1540-580X/2023.2308]
10. Richards CA, Doorenbos AZ. Intercultural competency development of health professions students during study abroad in India. J Nurs Educ Pract 2016 Dec;6(12):89-98. [doi: 10.5430/jnep.v6n12p89] [Medline: 29129984]
11. Dalia Y, Milam EC, Rieder EA. Art in medical education: a review. J Grad Med Educ 2020 Dec;12(6):686-695. [doi: 10.4300/JGME-D-20-00093.1] [Medline: 33391592]
12. Slavin R, Williams R, Zimmermann C. Activating the Art Museum: Designing Experiences for the Health Professions: American Alliance of Museums; 2023. URL: https://research.ebsco.com/linkprocessor/plink?id=7968f0f5-891d-33a5-891b-dfc77ebaac4a [accessed 2025-08-26]
13. Cioppi E, Gutiérrez NG, Lawrence E, et al. University Museums from Home: observations on responses to the impact of COVID-19. Univ Mus Collect J 2020;12(2):138-151 [FREE Full text]
14. Siri A. Living heritage educational experiences in a pandemic scenario. The case study of the ethnomedicine museum A. Scarpa. In: Kayapinar U, editor. Teacher Education: New Perspectives: BoD – Books on Demand; 2021:119-138. [doi: 10.5772/intechopen.96399]
15. Srivastava AA, Cohen S, Hailey D, Khoshbin S, Katz JT, Ganske IM. Training the eye, virtually: adapting an art in medicine curriculum for on-line learning. SN Soc Sci 2022;2(8):158. [doi: 10.1007/s43545-022-00442-4] [Medline: 35971453]
16. Ambrose M, Murray L, Handoyo NE, Tunggal D, Cooling N. Learning global health: a pilot study of an online collaborative intercultural peer group activity involving medical students in Australia and Indonesia. BMC Med Educ 2017 Jan 13;17(1):10. [doi: 10.1186/s12909-016-0851-6] [Medline: 28086875]
17. Sharp A, Thomson L, Chatterjee HJ, Hannan L. The value of object-based learning within and between higher education disciplines. In: Chatterjee HJ, Hannan L, editors. Engaging the Senses: Object-Based Learning in Higher Education: Routledge; 2015:97-116 URL: https://www.taylorfrancis.com/chapters/edit/10.4324/9781315579641-8/value-object-based-learning-within-higher-education-disciplines-arabella-sharp-linda-thomson-helen-chatterjee-leonie-hannan [accessed 2026-09-22]




XSL•FO
RenderX

18. Kador T. Object-Based Learning: Exploring Museums and Collections in Education: UCL Press; 2025. [doi: 10.14324/111.9781787354296]
19. Hackett S, Dawson M, Janssen J, van Tartwijk J. Defining Collaborative Online International Learning (COIL) and distinguishing it from virtual exchange. TechTrends 2024 Nov;68(6):1078-1094. [doi: 10.1007/s11528-024-01000-w]
20. Rubin J, Guth S. The Guide to COIL Virtual Exchange: Implementing, Growing, and Sustaining Collaborative Online International Learning: Routledge; 2023. [doi: 10.4324/9781003447832]
21. Hackett S, Janssen J, Beach P, Perreault M, Beelen J, van Tartwijk J. The effectiveness of Collaborative Online International Learning (COIL) on intercultural competence development in higher education. Int J Educ Technol High Educ 2023;20(1):5. [doi: 10.1186/s41239-022-00373-3] [Medline: 36713634]
22. Bhabha HK. The Location of Culture: Routledge; 2004. URL: https://research.ebsco.com/linkprocessor/plink?id=3e524577-3272-3f6c-b22e-93104ee2d47d [accessed 2025-08-28]
23. Lave J, Wenger E. Situated Learning: Legitimate Peripheral Participation: Cambridge University Press; 1991. [doi: 10.1017/CBO9780511815355]
24. Gaunt H, Chng HT, Shea R, et al. A new initiative in synchronous international learning: a Health Humanities Global Classroom. ASCILITE Publ 2024 Nov;23:601-606. [doi: 10.14742/apubs.2024.1412]
25. Anderson LW, Krathwohl DR. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Longman; 2001.
26. Fisher CD. Padlet: an online tool for learner engagement and collaboration. Acad Manag Learn Educ 2017;16(1):163-165. [doi: 10.5465/amle.2017.0055]
27. Liamputtong P. Focus Group Methodology: Principles and Practice: SAGE Publications Ltd; 2011. [doi: 10.4135/9781473957657]
28. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2006 Mar;5(1):80-92. [doi: 10.1177/160940690600500107]
29. Kleinman A. The Illness Narratives: Suffering, Healing, and the Human Condition: Basic Books; 1988. URL: https://research.ebsco.com/linkprocessor/plink?id=799f7a80-d6c7-3cdc-a260-fe79814a8089 [accessed 2025-08-28]
30. Tervalon M, Murray-García J. Cultural humility versus cultural competence: a critical distinction in defining physician training outcomes in multicultural education. J Health Care Poor Underserved 1998 May;9(2):117-125. [doi: 10.1353/hpu.2010.0233] [Medline: 10073197]
31. Hook JN, Davis DE, Owen J, Worthington EL, Utsey SO. Cultural humility: measuring openness to culturally diverse clients. J Couns Psychol 2013 Jul;60(3):353-366. [doi: 10.1037/a0032595] [Medline: 23647387]
32. Bora S, Staal W, van Hoof J, Jacobs B, Visnapuu-Bernadt P, Deschamps P. Online cross-cultural dialogue: a pilot for transcultural learning in postgraduate psychiatry training. Acad Psychiatry 2025 Apr;49(2):152-156. [doi: 10.1007/s40596-024-02109-z] [Medline: 39707106]
33. Hafferty FW. Beyond curriculum reform: confronting medicine's hidden curriculum. Acad Med 1998 Apr;73(4):403-407. [doi: 10.1097/00001888-199804000-00013] [Medline: 9580717]


## Abbreviations

**COIL:** collaborative online international learning
**HHGC:** Health Humanities Global Classroom
**HKU:** University of Hong Kong
**MB-HE:** museum-based health education
**NUS:** National University of Singapore
**OBL:** object-based learning
**UoM:** University of Melbourne


*Edited by M Balcarras; submitted 20.Sep.2025; peer-reviewed by G Shankar Ganesh, M Arsenault, MS Shafi; revised version received 11.Sep.2026; accepted 11.Sep.2026; published 28.Oct.2026.*

*Please cite as:*
*Gaunt HM, Cham KM, Knothe F, Lim CH, Shea R, Chng HT*
*Impact of a New Virtual Health Humanities Approach for Collaborative Online International Learning: Mixed Methods Study*
*Interact J Med Res 2026;15:e84474*
*URL: https://www.i-jmr.org/2026/1/e84474*
*doi:10.2196/84474*




XSL•FO
RenderX

© Heather Meredith Gaunt, Kwang Meng Cham, Florian Knothe, Chee Han Lim, Rosie Shea, Hui Ting Chng. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 28.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Classroom Integration of AI-Simulated Patients and Automated Feedback in Doctoral Psychotherapy Training: Longitudinal Mixed Methods Study

Ellen Converse[1*], MS; Ryan Louie[2*], PhD; Debra Safer[1], MD; Juan Pablo Pacheco[2], BS; William Fang[2], BS; Jamie Kent[1], PhD; Bruce Arnow[3], PhD; Diyi Yang[2], PhD

[1]PAU-Stanford PsyD Consortium, University Palo Alto, 1791 Arastradero Road, Palo Alto, CA, United States

[2]Department of Computer Science, Stanford University, 353 Jane Stanford Way, Stanford, CA, United States

[3]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, 401 Quarry Road, Stanford, CA, United States

[*]these authors contributed equally

**Corresponding Author:**
Ellen Converse, MS
PAU-Stanford PsyD Consortium, University Palo Alto, 1791 Arastradero Road, Palo Alto, CA, United States

## Abstract

**Background:** Psychotherapy education increasingly emphasizes technology-assisted methods that extend deliberate practice beyond human supervision. AI has enabled automated skill assessment and patient simulations, yet few studies have examined how students experience such tools in real-world settings alongside established practice modalities.

**Objective:** We aimed to evaluate students' experiences with AI-augmented deliberate practice embedded in an existing psychotherapy course sequence. Secondary aims included identifying implementation barriers, ethical and pedagogical considerations, and factors influencing students' acceptance of AI-assisted learning.

**Methods:** Participants were 29 first-year clinical psychology doctoral students in a 2-quarter psychotherapy sequence at an American Psychological Association–accredited program. In the fall course, students completed weekly peer role-plays and could optionally submit transcripts to CARE for AI feedback to supplement peer feedback. In the winter course, students completed required weekly deliberate practice through Skillsetter video vignettes and could optionally practice with CARE's voice-based AI-simulated patients, receiving AI and instructor feedback afterward. Surveys with Likert-type and free-response items were voluntary in the fall and required in the winter; 9 and 25 students completed them, respectively. Analyses included descriptive statistics, matched-pairs testing, and thematic analysis.

**Results:** In the fall course (n=9), where privacy concerns and survey nonresponse limited data on AI feedback, students rated peer feedback higher than AI feedback on being constructive and helpful (median 3, IQR 3-3 vs median 2, IQR 2-2; *P*=.02) and recognizing areas to improve (mean 5.22, SD 0.83 vs mean 4.33, SD 0.87; *P*=.046), citing transcription errors and a narrow focus on empathy while valuing AI's alternative phrasing. In the winter course (n=25), Skillsetter was rated higher than CARE only for learning from mistakes (mean 4.52, SD 2.00 vs mean 3.20, SD 2.04; *P*=.03); immersion, spontaneous responding, communication beyond words, and alliance building did not differ significantly. Students valued CARE for interactive practice not possible with vignettes but reported undetected vocal pauses (20/25, 80%), insufficient realism (14/25, 56%), and limited emotional expression (13/25, 52%), and raised concerns about AI replacing peer support (22/25, 88%) and peer interaction (20/25, 80%). Asked about AI broadly, students expressed mixed views, citing concerns about supplanted relational dynamics, nonconsensual data use, and generative AI's environmental impacts.

**Conclusions:** Embedding AI-augmented deliberate practice into an existing course sequence was feasible, but students' experiences favored established methods: they preferred human over AI feedback and described video vignettes as more realistic than voice-based AI interactions, while valuing AI-simulated patients for interactive practice other modalities could not provide. Implementation barriers included feedback quality, transcription accuracy, and instructor time required to supplement AI feedback. Perceptions of AI's instructional utility appeared contingent on students' trust in the technology, suggesting future deployments should weigh a sample's attitudes toward AI and its value relative to existing training resources.

***(Interact J Med Res 2026;15:e93199)*** doi:10.2196/93199

**KEYWORDS**

artificial intelligence; psychotherapy; deliberate practice; classroom deployment; AI-simulated patients



XSL•FO
RenderX

# Introduction

## Background

Psychotherapy is a structured clinical intervention wherein trained professionals use established psychological methods, primarily through dialogue and relational engagement, to alleviate distress, reduce psychiatric symptoms, and promote psychological functioning and well-being [1,2]. Psychotherapy is a collaborative process in which the therapeutic alliance, defined as the emotional bond between the therapist and patient, as well as agreement on the goals of therapy and the tasks required to achieve the therapeutic goals [3], is consistently associated with clinical improvement [4].

Therapeutic communication skills cannot be fully developed through lectures or reading alone; instead, they are cultivated through experiential practice, structured feedback, and critical reflection. Accordingly, traditional psychotherapy training emphasizes didactic instruction (structured, instructor-led teaching methods, such as lectures, seminars, or presentations that convey foundational theories and practical guidelines) alongside supervised clinical practice, role-play, and reflective exercises. Collectively, these methods facilitate real-time interaction and iterative refinement of core therapeutic skills [5].

Training psychotherapists, therefore, presents a distinctive pedagogical challenge: students must master not only theoretical and technical knowledge but also deeply interpersonal competencies, including empathic attunement, alliance building, and moment-to-moment responsiveness. A small yet steadily growing evidence base suggests that deliberate practice (DP) is an effective learning method that can outperform standard supervision and training [6]. Unlike didactic instruction, DP focuses on structured and intentional skill development through activities that include clearly defined goals, repeated practice, systematic opportunities for reflection, and real-time feedback from a supervisor or coach.

A common form of DP in graduate psychotherapy training contexts consists of a real-time session with a peer or clinical supervisor role-playing a simulated scenario and/or providing feedback following a rehearsal or therapy recording. Empirical evidence supports the effectiveness of DP in enhancing these skills. For example, learners who engaged in DP showed greater improvements in empathic expression compared to those who were trained using didactic methods alone [7]. Despite its potential benefits, the widespread implementation of DP in clinical psychology training programs is limited by the significant time and expertise required for 1-on-1, real-time supervision, and personalized feedback [6].

## Technological Innovations in Psychotherapy Training

Researchers have sought to enhance the accessibility and efficiency of psychotherapy training by developing technology-enabled platforms. Early efforts focused on structured tools for DP, such as video recording systems, which facilitate asynchronous supervision and reflective practice by enabling both students and instructors to review therapist-patient interactions outside of real-time sessions [8]. Similarly, video-response platforms such as Skillsetter (Theravue Inc) [9] offer structured opportunities for practice by presenting short patient vignettes and requiring trainees to record therapeutic responses to those vignettes, which can later be reviewed for feedback by peers or supervisors. These innovations have expanded opportunities for repeated skill rehearsal, supplementing traditional classroom and supervisory contexts.

Recent advances in large language models (LLMs) have further broadened this landscape, particularly through the development of AI-simulated patients that allow counselors to engage in practice dialogues within safe, controlled environments [10,11]. Parallel work has advanced AI feedback systems capable of automatically assessing microskills such as empathy, active listening, and motivational interviewing [12-16], while also generating alternative responses [8,17,18] and explanatory rationales [19,20].

Within the emerging literature, AI-augmented psychotherapy training systems show both potential and methodological limitations. Recent experimental laboratory studies suggest that such simulations, when paired with feedback, may promote the acquisition of patient-centered counseling behaviors [21,22]. Avatar-based simulated patients can effectively support motivational interviewing practice, garnering high levels of usability and satisfaction among both student and professional counselors [23]. More recently, AI-simulated patients have been assigned within clinical education settings such as psychotherapy [24,25] and medical education [26]. However, even the most promising implementations fall short of assessing how trainees engage with AI over extended periods alongside traditional coursework, supervision, and peer practice. Consequently, questions remain about how trainees engage with AI tools when integrated into coursework as part of ongoing real-world classroom settings rather than deployed as standalone exercises.

## The Current Investigation

Addressing the limitations of prior work, this study seeks to advance the understanding of how AI-augmented DP functions when directly embedded into psychotherapy training. As reviewed earlier, existing research on AI tools has primarily been restricted to controlled laboratory settings, short intervention windows, or isolated evaluations conducted outside of the classroom context. Such studies cannot fully address how AI interacts with real pedagogical structures, classroom dynamics, or established training methods such as peer role-play, supervised feedback, and video vignette exercises. Moreover, it remains unclear whether AI practice tools operate as meaningful complements to these traditional approaches, whether they inadvertently displace more valuable training modalities, or whether their integration reshapes learning in unanticipated ways.

To address these gaps, the present investigation examined students' experiences with CARE, a voice-based, AI-augmented DP platform on which users speak in real time with AI-simulated patients, mirroring the pacing and unpredictability of an actual patient encounter rather than clicking through scripted prompts, when deployed across 2 academic quarters within a doctoral-level psychotherapy curriculum. CARE was integrated alongside existing training methods, including peer role-plays



XSL•FO
RenderX

and the Skillsetter video-response platform, to enable structured comparison of AI-based approaches with established educational tools. The primary objective of the study was to evaluate the feasibility and perceived value of implementing CARE in this classroom context. Secondarily, we explored implementation barriers, ethical and pedagogical considerations, and factors influencing students' acceptance of and engagement with AI-assisted learning. Guided by these objectives, the study addressed 4 research questions (RQs):

- RQ1: Comparative value—From the perspective of students, how do 2 bundled DP packages, an AI-augmented package (CARE: voice-based, multiturn practice with AI feedback) and an established video-based package (Skillsetter: single-turn vignette practice with instructor feedback), compare in perceived instructional value, and what distinct affordances do students attribute to each package as delivered?
- RQ2: Usability and technical limitations—What factors about the AI-augmented package prototype (CARE) might enhance the pedagogical effectiveness of CARE within psychotherapy training?
- RQ3: Course integration influence on acceptability—How are utility and acceptability influenced by how the AI-augmented package (CARE) is embedded into the psychotherapy classroom, including issues of onboarding, communication, and strategies to preserve instructor-student and peer-to-peer interaction within the deployment?
- RQ4: Student beliefs about acceptance—(1) How do students perceive the legitimacy and acceptability of the AI-augmented package (CARE) in psychotherapy education, and (2) how has their exposure to AI in general influenced their views on its potential role in clinical training and practice?

Through its design, to our knowledge, the study provides one of the first longitudinal, classroom-based evaluations of an AI-augmented DP tool in psychotherapy education. By situating AI within authentic curricular structures rather than laboratory simulations, this work offers an ecologically valid assessment of both the pedagogical potential and the limitations of AI integration, thereby informing future directions for technology-mediated training in the mental health professions.

# Methods

## Ethical Considerations

The protocol for this psychotherapy classroom deployment study was reviewed by Stanford University Institutional Review Board and was assessed as nonmedical exempt (protocol ID 72546). This research was conducted as part of normal educational practice in an introductory psychotherapy course in a clinical psychology PsyD program at the Palo Alto University–Stanford PsyD Consortium. AI-augmented practice components were added to standard weekly assignments. The study included mid-class and end-class surveys to assess students' perceptions of AI feedback compared to peer feedback. Students received an information sheet explaining that while participation in course activities was expected as part of normal educational practice, they could opt out of having their data included in research analysis by emailing the protocol director. The instructor had no knowledge of which students opted out, ensuring no coercion or impact on grades. No compensation was provided as the AI-augmented assignments and surveys were integrated into normal course requirements.

To protect privacy and confidentiality, all data transmission was encrypted via HTTPS and stored on password-protected Amazon Web Services (AWS) servers accessible only to investigators listed on the protocol. Survey responses were linked to deidentified participant IDs through Qualtrics (Qualtrics International Inc). The research team committed that practice transcripts would not be publicly released or used to train autonomous AI therapy chatbots; we further describe in *Data Privacy Protection in CARE's System Architecture* section how we honor that commitment within our data and AI processing pipelines. Audio recordings and transcripts would be stored for up to 3 years to allow for the completion of all scientific analyses by the authors before deletion. Only deidentified survey data, aggregated usage statistics, and feedback ratings could be presented at scientific meetings or published in journals. The AI feedback system underwent safety validations prior to deployment, including guard models to filter inappropriate content.

## Participants

Participants were 29 first-year doctoral students enrolled in the abovementioned PsyD program. Of these participants, 4 students elected not to engage with CARE at any point during the study. The remaining 25 students participated in at least some component of the CARE activities and comprise the analytic sample reported here.

Students entered with heterogeneous training backgrounds: some held master's degrees and had accumulated significant supervised clinical hours, whereas others entered with bachelor's degrees supplemented by applied fieldwork or human services experience. This requirement for prior clinical experience differentiates PsyD training from many master's mental health programs and some PhD programs, which often emphasize research over early clinical immersion.

Demographic data (eg, age, gender, and ethnicity) were not collected, which limits analysis of subgroup variation.

## Pedagogical Context and Course Goals

The study was embedded within a 2-quarter sequence forming part of the core first-year PsyD curriculum. This sequence is designed to provide foundational training in psychotherapy competencies, with a structured progression from introductory relational skills to more advanced therapeutic intervention. The study was conducted over 16 weeks, spanning 2 consecutive courses (September 2024 to March 2025).

The first course study emphasized pantheoretical elements of psychotherapy, including basic listening skills, therapeutic alliance formation, therapist verbal response models and intentions, and consideration of cultural and ethical factors. The second course quarter built upon these foundations by targeting pantheoretical competencies outlined by the American Psychological Association (APA) Division 29 Task Force on



XSL•FO
RenderX

Psychotherapy Competence, such as alliance maintenance (task, goal, and bond), rupture repair, transference and countertransference, and termination processes.

### Historical Structure of the Learning Psychotherapy Course Series

In prior cohorts, the first course used weekly peer roleplays. Students alternated therapist and patient roles in 10-minute sessions, followed by mutual feedback and written self-reflections. Confidentiality agreements were required, and each student submitted recorded sessions that were periodically reviewed in small group meetings with instructors.

In the second course, prior cohorts engaged with Skillsetter's video-based DP platform. Assignments required demonstration of specific therapeutic skills (eg, empathy, cultural humility, and alliance repair). Instructors or teaching assistants (TAs) provided feedback on a weekly basis, and the quarter culminated in a live roleplay with an instructor or TA portraying a patient, requiring integration of multiple therapeutic competencies. This established format of peer roleplay in the first course and Skillsetter video-based DP in the second course provided both relationally grounded and structured opportunities for iterative skill refinement.

### Skillsetter Platform

Skillsetter is a web-based, video-based DP training platform designed to support practice in interpersonal and counseling skills [27]. It was developed to overcome logistical barriers commonly encountered in traditional psychotherapeutic training, such as the resource-intensive nature of instructor supervision and inconsistent feedback, by allowing learners to practice asynchronously, record responses to simulated prompts, and receive structured feedback. When using the platform, students submitted video-recorded responses to standardized patient vignettes performed by actors. Because feedback, rather than mere repetition, is central to DP, Skillsetter emphasizes scaffolded evaluation, either via instructor raters or automated scoring across domains.

The platform has achieved broad adoption across counseling, social work, and psychotherapy training programs, lending evidence of its acceptability and institutional integration. In the Learning Psychotherapy course series, feedback on Skillsetter practice was given by the course instructor and TAs. In applying it to this study, Skillsetter served as an established, scalable platform for implementing component-level practice (eg, empathy, reflections, and open questions) in a remote environment. This enabled learners to engage in deliberate rehearsal outside class hours while receiving targeted feedback.

### Integration of CARE

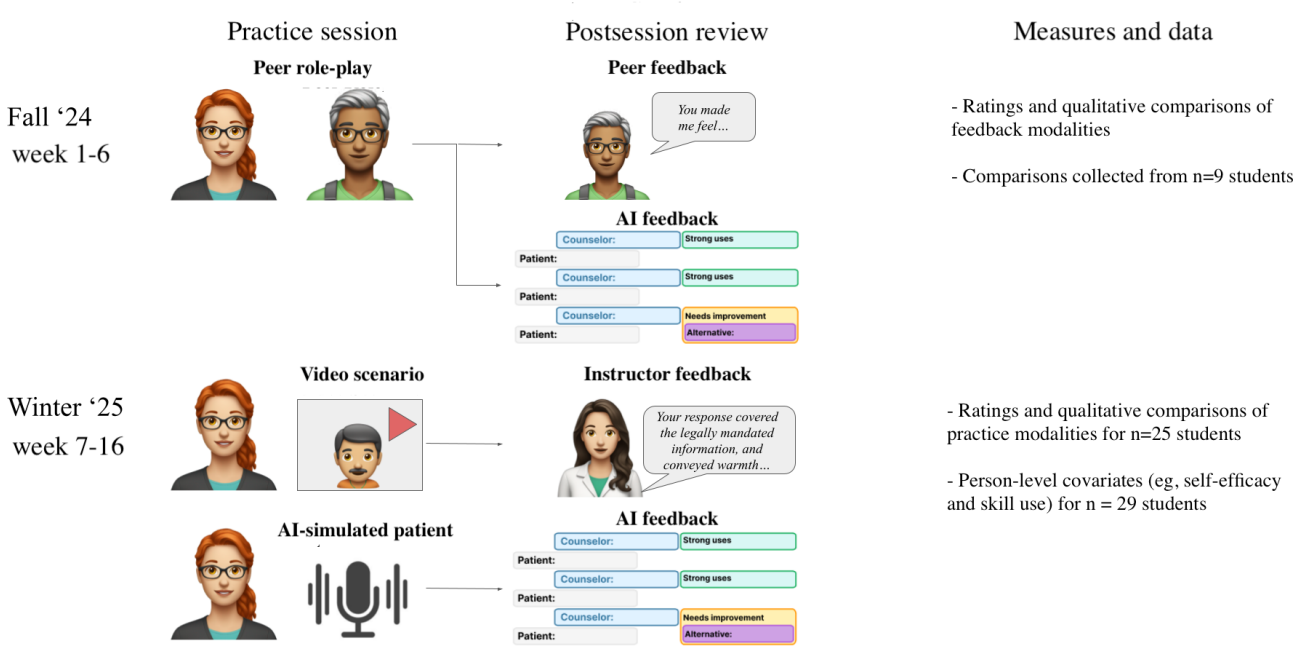
Building on this structure, this study introduced CARE, an AI-powered simulation and feedback platform, to examine its potential role alongside traditional DP (Figure 1). Our design maintained the integrity of the existing pedagogy while creating opportunities for systematic comparison.

In the first course pilot study in 2024, the instructional team prioritized interpersonal skill building and peer connection, continuing to rely on peer role-play as the primary experiential method. CARE was introduced as an optional feedback tool, allowing students to upload recorded peer sessions and receive automated feedback (Figure 2) on the use of 8 counseling skills (eg, empathy, validation, reflections, questions, suggestions, self-disclosure, session structure, and professionalism) derived from the helping skills and motivational interviewing models [28,29]. This integration provided an opportunity to explore the additive value of AI-generated feedback while preserving the emphasis on human relational practice in the early stages of training.



XSL•FO
RenderX

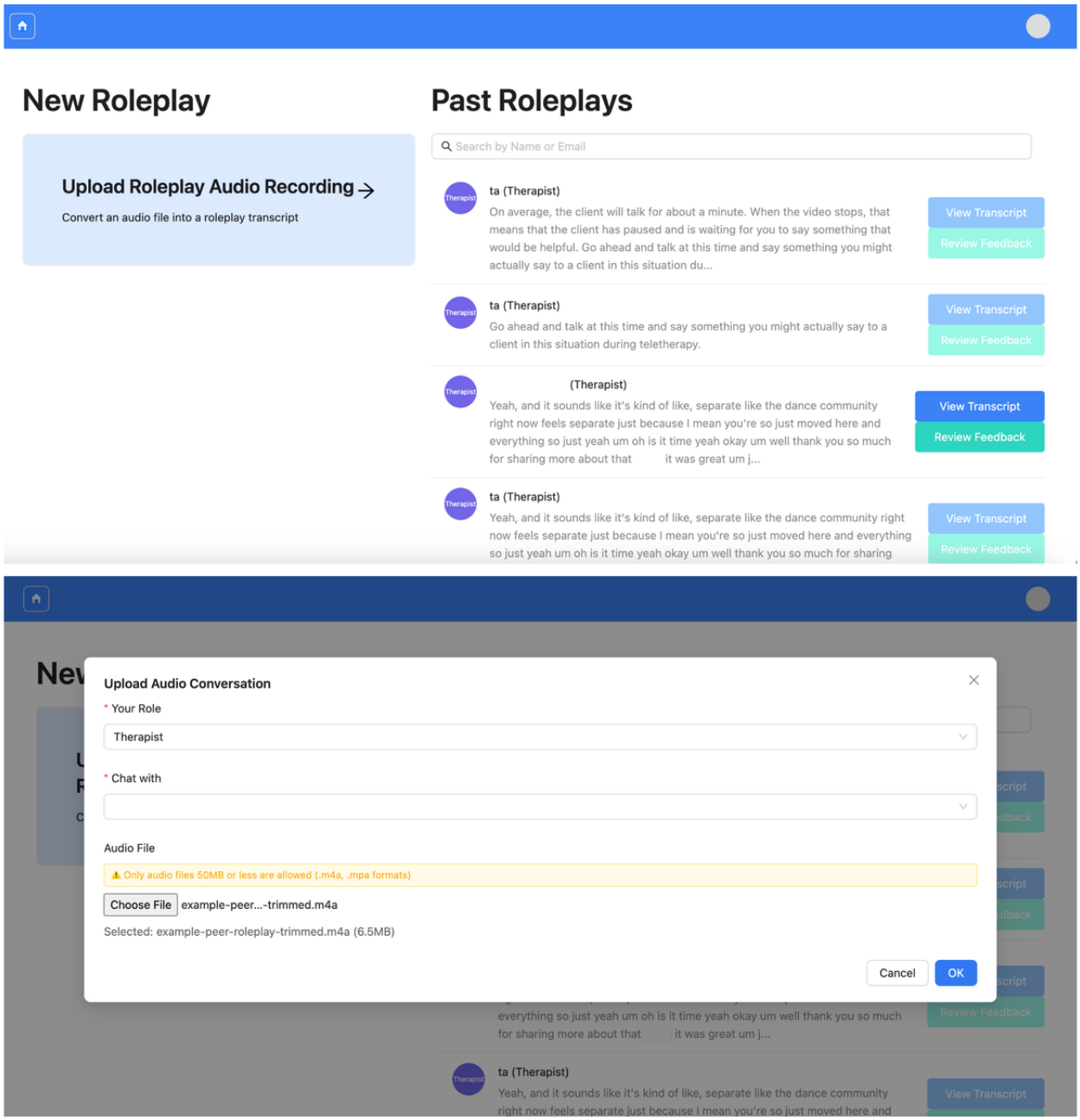
**Figure 1.** Overview of the integration of CARE's practice, feedback, and data collection into the 2 courses. The first course implementation lasted 6 weeks, and the second course lasted 10 weeks. During the first course (weeks 1 - 6), students engaged in peer role-play practice sessions followed by post-session review using either peer-provided or AI-generated feedback, with ratings and qualitative comparisons collected (n=9). During the second course (weeks 7 - 16), students completed practice using video-based scenarios and AI-simulated patients, followed by instructor or AI-generated feedback, with ratings, qualitative comparisons, and person-level covariates (eg, self-efficacy and skill use) collected (n=25). The figure illustrates the progression of practice modalities, feedback sources, and data collection across the 2 courses.






XSL•FO
RenderX

**Figure 2.** Workflow for AI feedback on peer role-plays in the first course. Students completed weekly peer-to-peer role-plays, alternating between therapist and patient roles, and recorded each session. Audio files uploaded to CARE were transcribed and diarized into speaker turns, and the platform generated feedback on the student's utterances as therapist. The content shown is illustrative example data from test accounts; no student data are displayed.




In the second course (CLIN 702), CARE was deployed more comprehensively in parallel with Skillsetter to enable structured comparison of 2 distinct DP modalities (Table 1). Students alternated between Skillsetter's single-turn vignette responses and CARE's multiturn, dynamic dialogues with AI-simulated patients. Both platforms were aligned with weekly curricular objectives, matched for duration and credit, and graded using consistent rubrics. This dual-modality design created an ecologically valid setting for evaluating the relative affordances and limitations of AI-augmented practice versus established methods.



XSL•FO
RenderX

**Table .** Comparison of the 2 bundled deliberate practice packages evaluated in course 2[a].

| Dimension | CARE: voice-based AI-augmented package | Skillsetter: video-based package |
|---|---|---|
| Practice format | Multiturn, dynamic spoken dialogue with an LLM[b]-driven AI-simulated patient (voice; GPT-4o Realtime API) | Single-turn video-vignette response: 1 recorded reply to a standardized patient video performed by an actor |
| Feedback source | Automated AI feedback on each utterance (fine-tuned Llama 2 13B "mentor"); supplemented midquarter by weekly written TA[c] feedback to 3 randomly selected students who completed a practice session | Human feedback from the course instructor and teaching assistants, using established rubric-based scoring |
| Number of sessions | 10 weeks, 2 to 4 simulated patients offered as options per week | 10 weeks, 1 simulated patient per week |
| Scenario type | AI-simulated patients spanning alliance ruptures, mandated reporting, and suicide risk/self-injury adapted from Skillsetter transcripts (9/12) and documented cases of sexually inappropriate behavior (3/12) | Standardized, actor-performed vignettes targeting specific skills (eg, empathy, cultural humility, and alliance repair) |
| Required versus optional | Initially required but was made optional in the second week. Students could complete additional Skillsetter practice instead of CARE | Established, required component of the course |
| Duration | Matched to Skillsetter for time and credit; weekly across weeks 7 to 16 (1 - 2 minutes per session) | Weekly across weeks 7 to 16 (1 - 2 minutes per session) |
| Grading | Rubrics matched to Skillsetter for evaluation criteria and grading weight | Rubric-based grading; quarter culminated in a live role-play via videoconference with an instructor or TA portraying the patient |

[a]Course 1 involved a different comparison (AI-augmented feedback and peer feedback on peer role-plays) and is described separately.

[b]LLM: large language model.

[c]TA: teaching assistant.

## Safety Procedures for Emotionally Sensitive Scenarios

CARE's scenario library included clinically challenging content, including alliance ruptures, mandated reporting, suicidal ideation and self-injury, and sexually inappropriate behavior toward clinicians. These topics were not introduced by the study. They are established components of the course curriculum, which addresses suicide risk and mandated reporting in week 7 and sexual attraction and boundary maintenance in week 8, and they reflect competencies specified by the APA Division 29 Task Force on Psychotherapy Competence [30]. Exposure to such material under supervision is a required element of doctoral clinical training, and these topics had been routinely taught in this course sequence with prior cohorts. CARE scenarios were mapped onto the existing weekly syllabus rather than developed independently, so students encountered simulated versions of material they were concurrently studying in class and practicing in Skillsetter.

Before deployment, each AI-simulated patient was reviewed by the TA and co-first author (EC), who had previously completed this course and held clinical training. Scenarios were additionally reviewed by the course instructor (JK), a licensed clinical psychologist and co-associate director of clinical training for the program, who confirmed their clinical appropriateness and alignment with the week's learning objectives before they were released to students. Review criteria were clinical appropriateness, ethical soundness, and alignment with the week's learning objectives. For scenarios involving trauma, self-harm, or suicidal ideation, review specifically assessed whether the AI-simulated patients supported rather than undermined the learning objective in the curriculum. Behavioral principles constraining each simulated patient were revised iteratively during this process. Nine of the 12 scenarios were adapted from existing Skillsetter transcripts already in curricular use; the remaining 3 were adapted from documented cases in the published literature [31].

Faculty oversight operated through the courses' existing structures rather than through study-specific monitoring. Students had established channels for raising concerns: in-class discussion, direct contact with the instructor or TA, and the program's existing student support structures. As the instructor and TA reviewed Skillsetter submissions on the same weekly topics, distress surfacing in either modality could be identified and addressed through the same channels.

We did not review CARE session transcripts in real time, and no automated system flagged student distress during practice. Safeguards were therefore preventive rather than responsive: scenario prompts were constrained in advance, and guard models filtered inappropriate simulated-patient output before students encountered it. No students reported clinically inappropriate or harmful content from the simulated patients. However, because we did not systematically review sessions, we cannot confirm that none occurred. We address this further in the "Limitations" section.



XSL•FO
RenderX

## CARE AI System

CARE was developed as a web-based training platform designed to support novice counselors in practicing counseling skills within simulated environments powered by LLMs. The platform enables trainees to practice psychotherapy skills through roleplay conversations either with their peers or with AI patients and to receive automated feedback on their therapeutic responses, generated by an AI "mentor." CARE builds on prior work in co-design with mental health experts to enhance the realism of AI-simulated patients [8,10,11,26] and on research advancing the fine-tuning of domain-specific LLMs to incorporate therapeutic knowledge capable of producing both evaluative feedback and alternative counselor responses [6,19,32-34].

From a technical standpoint, CARE was implemented as a web application with a Python Flask (Pallets) back end and a React (Meta) JavaScript front end, accessible to students through password-protected student accounts. In the first course, the CARE platform allowed students to upload an audio recording of a peer role-play conversation, which was automatically converted into a diarized transcript. The transcription process used the WhisperX system (open source) and the OpenAI API to help with LLM-assisted diarization between therapist and patient utterances. In the second course, students practiced a role-play conversation with AI-simulated patients. AI-simulated patients were powered by the GPT-4o Realtime API, with scenario-specific prompts defining patient background, conversational style, and behavioral principles [10]. The OpenAI Whisper model was also used to transcribe the audio dialogue into text-based utterances. This infrastructure enabled integration into classroom deployment studies, allowing systematic investigation of CARE as a pedagogical supplement to established DP methods.

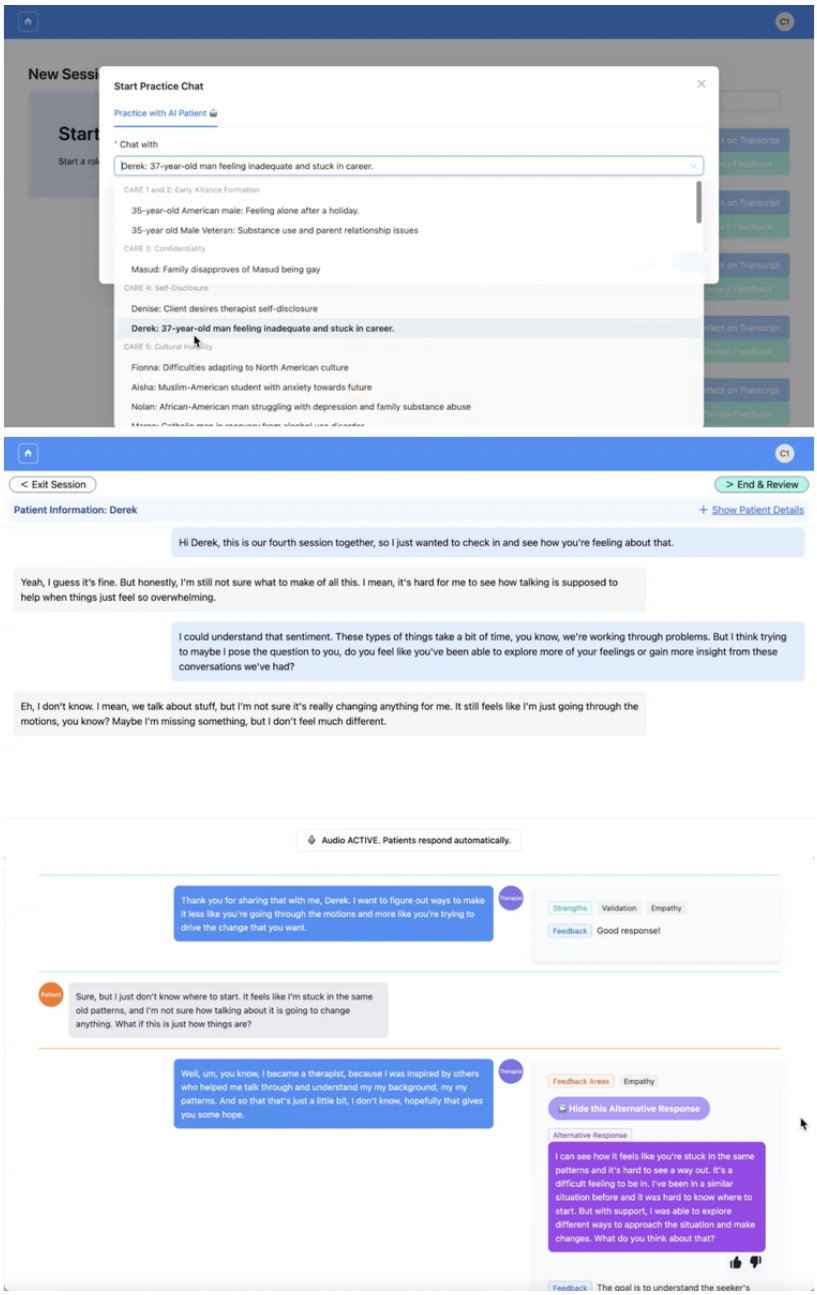
Within the version of CARE deployed to the second course, practice was organized around a library of patient scenarios, each defined by a brief background vignette (eg, "Young adult with family issues: low mood and self-esteem"). These intentionally limited descriptions are designed to approximate authentic counseling encounters, in which the clinician must simultaneously build rapport, elicit patient context, and apply therapeutic skills (Figure 3). For example, a novice peer counselor might initiate a practice session with an AI-simulated patient modeled as a 35-year-old veteran struggling to reconnect with his children amid legal and interpersonal challenges. The dialogue proceeds turn by turn, requiring the trainee to apply core counseling skills such as empathy, reflection, and problem exploration, while managing relational complexity as the conversation unfolds.

The platform provided students with AI-generated feedback on each of their utterances at the end of the session. CARE's feedback implements an existing method for generating feedback that fine-tunes a Llama 2 13B (Meta) parameter model on an expert-annotated dataset covering psychotherapy-relevant skill assessment [19]. Feedback focuses on strengths, areas for improvement, feedback explanation, and a suggested alternative response that addresses any issues.



XSL•FO
RenderX

**Figure 3.** Weekly practice with AI-simulated patients in the second course. CARE offered students a set of voice-based AI-simulated patient scenarios each week, mapped onto that week's learning goals (eg, therapist self-disclosure and cultural humility). Following each practice session, the platform generated feedback on the student's utterances as therapist. The content shown is illustrative example data from test accounts; no student data are displayed.




## Data Privacy Protection in CARE's System Architecture

CARE's system architecture was designed to protect the privacy of student audio recordings and voice data and ensure they were not used for training AI models. In places where OpenAI APIs were used, the data policy was set to opt out of using any input or output data for AI model training; the OpenAI data retention policy further states that only abuse monitoring logs are retained for a maximum of 30 days.

For the pipeline that transcribed student peer role-play audio (course 1), uploaded audio files were saved onto the private AWS EC2 instance; audio was transported to an ephemeral, serverless graphics processing unit (GPU) service that started the WhisperX transcription and diarization function; the transcript was given to a GPT-4o-assisted diarization pipeline to identify therapist and patient roles; finally, the therapist/patient messages were saved to a SQLite database on the AWS EC2 instance. In this way, long-term data are only retained on our private AWS EC2 instance and are governed by our maximum 3-year retention policy to support scientific analysis.

The pipeline to practice with voice-based AI-simulated patients (course 2) is fully handled by the gpt-realtime API with authentication to OpenAI handled through ephemeral client API keys issued by the server. The gpt-realtime API end point responds to audio events and uses the OpenAI Whisper API to transcribe the audio transcript. Our OpenAI data privacy policy



XSL•FO
RenderX

restricts long-term data retention and training on data sent to the API.

The feedback generation pipeline relies on an ephemeral, serverless GPU service that loads the fine-tuned Llama 2 13B model weights and processes the utterances in the transcript. The generated feedback is saved into the SQLite database. Thus, transcript and feedback data are only retained on the private AWS EC2 instance and are governed by our maximum 3-year data retention policy to allow for scientific analyses.

## Integration of CARE Into the Learning Psychotherapy Course Series

### *Overview*

The Learning Psychotherapy sequence spanned 2 quarters, 16 weeks total, and was structured to provide developmentally sequenced training in core psychotherapy competencies. The first course implementation spanned 6 weeks, whereas the second course implementation extended across 10 weeks. The study team began the first course in week 4 of the first course instead of week 1 to allow more time for finalizing implementation procedures.

The first course 2024 pilot, described in more detail in the following section, served as an initial, formative phase of the study, designed to explore the educational utility of AI-generated feedback with traditional peer feedback following role-play exercises. This phase was crucial in identifying themes that shaped the subsequent second course study, including the relative value of different feedback modalities (RQ1), early design limitations (RQ2), challenges of classroom integration (RQ3), and students' broader attitudes about the role of AI in psychotherapy training and daily life (RQ4).

### *First Course: Foundational Skills and Peer Role-Play*

The first course provides a foundation for training beginning psychotherapists in the crucial elements of psychotherapy. Across weeks 1 to 7, instruction progressed via the gradual introduction of the core listening skills and verbal response modes, with particular attention in week 4 to how sexual, gender, and racial diversity can impact the therapeutic alliance. Weeks 8 to 9 introduced principles and strategies from motivational interviewing, with the course concluding in week 10 with a focus on ethics, professional boundaries, and therapist self-care.

Students engaged in 10-minute peer role-plays each week, alternating between therapist and patient roles, with assigned therapeutic skills tailored to that week's topic. Recordings of these peer role-plays were uploaded to CARE, which generated automated feedback across 8 core counseling skill domains derived from the helping skills and motivational interviewing models. This process enabled the instructional team to evaluate student perceptions of AI-generated feedback as a supplement to traditional experiential exercises, while maintaining the central role of in-person peer practice in early clinical training.

### *Second Course: Comparative Deployment of Skillsetter and CARE*

The second course introduced students to the principles of DP, values-based care, and basic therapeutic skills. Instruction progressed from the concept of DP (week 1) to the therapeutic alliance, drawing on Bordin's task-goal-bond framework (week 2) [35], before moving to therapist self-disclosure, cultural humility, and the influence of power and identity within therapeutic relationships (weeks 3 and 4). The middle phase emphasized countertransference, resistance (week 5), alliance ruptures and repair (week 6), and clinical challenges such as mandated reporting and suicide risk (week 7). The sequence concluded with advanced relational concerns, including sexual attraction and boundary maintenance (week 8).

The course was designed to directly compare 2 distinct DP modalities: Skillsetter and CARE. This parallel design created a side-by-side comparison of vignette-based practice and conversational AI simulations, with both modalities aligned to weekly curricular goals and equivalent in duration, evaluation criteria, and grading weight.

To integrate CARE into the curriculum, the study team systematically mapped AI-simulated patients onto weekly course topics at week 4, enabling students to engage with simulated clinical encounters that corresponded directly with concepts under instruction. Students practiced with 12 AI-simulated patients across the quarter, 9 (75%) of which were adapted from Skillsetter transcripts and 3 (25%) adapted from documented examples of sexually inappropriate behaviors toward clinicians reported in prior empirical work [31] (Multimedia Appendix 1). CARE's scenarios encompassed alliance ruptures, mandated reporting, and crisis management related to suicidal ideation and self-injury, thereby anchoring AI practice in clinically relevant challenges.

Prior to classroom deployment, CARE's AI-simulated patients underwent an iterative review process conducted by the TA to ensure clinical appropriateness, ethical safeguards, and pedagogical alignment. Each AI-simulated patient was programmed to follow structured behavioral principles, with particular emphasis placed on scenarios involving emotionally sensitive content areas (eg, trauma and self-harm) to verify that AI-generated responses were trauma informed, ethically sound, and supportive of therapeutic learning objectives. Additional realism was introduced through varied voice models designed to convey emotional depth and natural conversational flow. Some identified limitations, including emotional flatness in voice-based responses, were addressed by adding principles to the voice-based AI-simulated patient's prompt (eg, "Mention past attempts to please family that failed, with resigned sighs, and even breaths where the nose is runny as in crying").

### *Measurements*

### *Surveys*

#### First Course

Students completed a midcourse survey in the first course, halfway through the introduction of CARE, and again at the end of the course. The 37-item questionnaire was specifically designed by the investigators for this study to gather insights into the challenges and obstacles students faced while using CARE, as well as their experiences with peer feedback and AI feedback on their peer roleplays. The full set of 37 items for the



XSL•FO
RenderX

first course study is provided in Multimedia Appendices 2 and 3.

### Second Course

To evaluate the comparative value of CARE and Skillsetter, students completed a 39-item investigator-designed survey at the end of week 16 in the second course. Items were adapted from national and professional surveys where possible [36,37]. Students rated their experiences across domains using 7-point Likert scales (1=strongly disagree to 7=strongly agree) and provided open-ended responses. Example items by domain are shown in Table 2. The full set of 39 items for the second course is provided in Multimedia Appendix 4.

**Table .** Example survey items from the web-based survey administered in the second course.

| Domain | Example items |
|---|---|
| Comparative value (RQ[a]1) | • "The way I could use [CARE/Skillsetter] allowed me to make mistakes I could learn from."<br>• "After conversing with the patients in [CARE/Skillsetter], I feel more equipped to build and maintain a therapeutic alliance." |
| AI feedback (RQ1) | • "Receiving AI feedback shortly after completing CARE practice enhanced my learning experience."<br>• "I know when to ignore AI feedback when it is incorrect or does not make sense." |
| Self-reflection | • "I found the self-reflections after a [CARE/Skillsetter] session valuable for understanding how to improve." |
| Limitations and design (RQ2) | • "My concerns about CARE AI's patients include: voices sound robotic; transcripts are inaccurate."<br>• "My concerns about CARE AI's feedback include: wish it tracked progress over time." |
| Course integration (RQ3) | • "Would you recommend CARE AI to next year's PsyD cohort as a deliberate practice option?" |
| Attitudes about AI (RQ4) | • "Overall, would you say the increased use of AI in daily life makes you feel… (concerned vs. excited)?"<br>• "Do you think AI in the classroom would lead to better or worse helping skills for therapists?" |

[a]RQ: research question.

## Instrument Development

Both surveys were investigator designed for this deployment, as no validated instrument exists for assessing student experiences with AI-augmented DP in psychotherapy training. Items were developed by the research team, which combined expertise in psychotherapy training and in AI systems for clinical education; the instructor and coinvestigators drew on prior experience teaching this course sequence to anticipate likely obstacles, including discomfort with recorded practice, concerns about data use, and skepticism about automated feedback on relational skills. Closed-ended response options for items on limitations and concerns were generated from these anticipated obstacles and from issues students raised during the first course. Items on general AI attitudes were adapted where possible from national and professional surveys [36,37]. The first-course survey served as a formative pilot, and its results informed revision of the second-course instrument.

No formal psychometric evaluation was conducted. Items were not subjected to cognitive interviewing, factor analysis, or reliability testing, and we report each item individually rather than computing scale scores. Quantitative findings should therefore be read as structured summaries of student reports rather than as validated measurements.

## Data Analyses

### *Quantitative Analyses*

We conducted nonparametric paired hypothesis tests (2-tailed Wilcoxon signed-rank) to evaluate students' comparative perceptions of CARE versus the 2 traditional DP modalities: peer role-play feedback and Skillsetter (RQ1). Our investigation concerns different dimensions of student perceptions (eg, comfort in receiving feedback from AI vs peers and immersiveness of voice-based AI and video-based vignette practice) that we consider equally and independently rather than as surrogate indicators of an omnibus finding; therefore, each dimension defines an individual statistical hypothesis tested at an  of .05 level without correction [38,39]. There were no missing data to account for in these paired survey responses. Statistical testing was conducted using the SPSS package (version 29; IBM Corp). Descriptive statistics are reported for the remaining quantitative items, covering perceived limitations and design needs of CARE (RQ2), perceptions of course integration (RQ3), and attitudes toward AI (RQ4).

### *Qualitative Analyses*

Open-ended survey responses were analyzed using reflexive thematic analysis following Braun and Clarke [40,41]. Coding



XSL•FO
RenderX

was hybrid: codes were generated inductively from the response data, while the 4 research questions, which structured the survey instrument and therefore the domains in which students responded, oriented the analysis. Two authors (EC and RL) independently read the full response set for familiarization and independently coded all responses. They then met over 3 rounds to compare codes, discuss divergent readings, and develop shared interpretations of the data. Rather than resolving differences by adjudication or by calculating interrater reliability, we treated points of divergence as analytically generative, working through them in discussion until we arrived at interpretations both coders found defensible. This is consistent with the assumptions of reflexive thematic analysis, which positions researcher subjectivity as a resource rather than a source of bias to be controlled [42]. Candidate themes were then developed from the coded material, reviewed against the coded extracts and the full dataset, refined, and named.

### *Classroom Memos*

Throughout the deployment, the TA and the co-first author kept informal written notes documenting classroom interactions, student questions raised during class, and technical problems reported in real time. Notes were written within 24 hours of each class session and were not systematically coded or analyzed as a data source. They served 2 limited purposes: to document implementation events that would otherwise go unrecorded (eg, the midquarter decision in course 2 to add supplemental TA feedback in response to student concerns) and to help the analytic team interpret survey responses that referenced specific classroom occurrences. We do not draw findings from the memos and do not treat them as qualitative data; where they inform the narrative, this is stated explicitly in the text.

### *Integration of Quantitative and Qualitative Findings*

We used a convergent design in which both strands were collected concurrently within the same surveys and analyzed separately, then integrated at the interpretation stage [43]. Integration proceeded by research question: for each RQ, we compared the direction and strength of the quantitative findings against the corresponding themes, noting points of convergence, divergence, and expansion. Convergence and divergence are reported explicitly in the “Results” section and revisited in the “Discussion” section.

### *Reflexivity*

Our team’s composition shaped both the deployment and its interpretation. The team includes computer scientists who developed CARE and its underlying models and clinicians who teach in the training program where the study took place. As developers, we had an investment in CARE functioning well. As instructors and program faculty, we held evaluative authority over the students whose experiences we report. Both students’ willingness to voice criticism and our readiness to hear it were conditioned by these relationships.

Several steps mitigated these risks. Research participation was decoupled from course participation; CARE was made optional in the first course as a supplement to peer feedback on students’ peer role-plays. In the second course, CARE was initially required; however, in the second week of the course, the instructional team made participation with CARE optional due to student concerns, and students could opt to complete weekly DP assignments using the Skillsetter video-based package only.

In coding and interpretation, the 2 coders included 1 coauthor without a role in developing CARE and 1 without. This positioning also bounds what the paper claims. We report students’ perceptions rather than an efficacy evaluation of CARE, and we treat students’ reluctance and ethical objections as findings in their own right rather than as barriers to be overcome.

## Results

### Overview

This section outlines the findings from the 16-week deployment, structured to provide a clear progression of insights. We begin with the first course 2024 pilot study, which served as a preliminary investigation into students’ perceived value of AI feedback as a supplement to peer feedback. We then move to the more extensive second course study, which builds upon the first course to explore AI-simulated patients in greater depth. This ordering reflects the evolving nature of the research and the increasing complexity of the questions addressed.

### First Course 2024 Pilot: AI Versus Peer Feedback

The first course was a formative phase of the study in which we sought to understand students’ experiences of AI-generated feedback and traditional peer feedback following peer role-play exercises. Of the 29 enrolled students, 23 uploaded at least one peer role-play transcript to CARE and 22 received AI feedback on at least one transcript (Table 3). All findings reported below reflect survey respondents only. The survey was voluntary in this course, and 14 (48%) students completed it. The following results are based on 14 survey respondents, 5 of whom did not use CARE by choice or due to technical issues, leaving 9 students who reported on the experience of AI versus traditional peer feedback.



XSL•FO
RenderX

**Table .** Prespecified implementation (feasibility) indicators by course.

| Indicator | Definition | Course 1 (CARE optional) | Course 2 (CARE mandatory in the first week) |
|---|---|---|---|
| Adoption | Percentage of enrolled who used CARE ≥1 time | 22/29 (76%) received CARE feedback on their peer role-play | 25/29 (86%) used CARE's AI-simulated patient (mandatory) |
| Reach | Percentage of enrolled providing any research data | 14/29 (48%; survey 1); 9/29 (31%; survey 2) | 25/29 (86%) |
| Completion rate[a] | Percentage of possible assignments completed with CARE with ≥4 dialogue exchanges | 73/174 (42%) | 304/319 (95%) |
| AI feedback review rate | Percentage of eligible transcripts that received AI feedback | 63/73 (86%) | 248/304 (82%) |
| Completion per student, mean (SD) | Sessions/uploads per student across the course | 3.17 (2.15) | 10.13 (4.69) |
| AI feedback reviews per student, mean (SD) | Sessions with feedback per student across the course | 2.74 (2.14) | 8.27 (3.91) |
| Assignment substitution | Whether CARE replaced or supplemented existing work | Optional add-on to peer role-play (no substitution) | Additional Skillsetter |
| Technical transcription failures | Rate of failed/unusable sessions due to transcription errors, including transcription to non-Latin scripts | Logs: non-Latin script in transcription, 0;<br>transcript repetitions, 2<br>Survey:<br>1 user reported incorrect diarization | Logs: non-Latin script in transcription, 4; missing speaker, 93;<br>AI patient stopped replying, 44 (midconversation, 15; start of conversation, 29)<br>Survey:<br>Voice-pause detection, 20/25 (80%) users; transcript inaccuracy, 10/25 (40%) users |
| Time required | Setup and per-session time for students/staff | 10 minutes | 10 minutes |
| TA workload | Staff hours for setup, scenario review, and feedback | Scenario review and troubleshooting: 1 - 2 hours per week | Weekly written feedback to 3 - 5 students and review: 1 - 2 hours per week |

[a]Course 1 completion rate assumes 6 chances to upload weekly peer role-plays to CARE×29 participating students; the rate is deflated because CARE use was optional and combines high-usage and no-usage students. Course 2 completion rate assumes 319 required assignments (11 required assignments×29 participating students); the rate is inflated because some students completed multiple sessions per week.

### *RQ1: AI Feedback Versus Peer Role-Play*

Quantitative comparisons from the first course pilot (Table 4) show that students rated peer feedback significantly higher than AI feedback along 2 of the 3 dimensions. Specifically, peer partners' feedback was rated as being more constructive and helpful to a great extent versus a moderate extent (median 3, IQR 3-3 vs median 2, IQR 2-2 on a 0 - 4 scale; $P$=.02). Students' peer partner feedback allowed them to better recognize areas to improve versus CARE with Likert ratings differing between moderately agree and slightly agree (median 5, IQR 5-6 vs median 4, IQR 4-5; $P$=.046). No statistically significant difference was found with regard to their comfort receiving feedback; the median student felt comfortable receiving feedback from peers and CARE's AI system (median 6, IQR 6-6 vs median 6, IQR 5-6 on a 1 - 6 scale). Nonetheless, peer partner feedback had higher mean and lower variance than CARE in terms of comfort (mean 5.88, SD 0.33 vs mean 5.44, SD 0.73), which indicates that a subset of students felt less comfortable with AI feedback.



XSL•FO
RenderX

**Table .** Comparing peer role-play practice followed by AI feedback (CARE) versus peer feedback (peer partner) rated by students in the first course (n=9).

| Question and modality | Mean (SD) | Median (IQR) | *P* value[a] |
|---|---|---|---|
| Constructive and helpful[b] | | | .02 |
| CARE | 2 (0.707) | 2 (2-2) | |
| Peer partner | 3 (0.707) | 3 (3-3) | |
| Comfort receiving feedback[c] | | | .10 |
| CARE | 5.444 (0.726) | 6 (5-6) | |
| Peer partner | 5.889 (0.333) | 6 (6-6) | |
| Recognize areas to improve[c] | | | .046 |
| CARE | 4.333 (0.866) | 4 (4-5) | |
| Peer partner | 5.222 (0.833) | 5 (5-6) | |

[a]*P* values are from Wilcoxon signed-rank tests comparing CARE and peer partner responses.

[b]Rated on a 0-4 scale (0=not at all, 1=to a small extent, 2=to a moderate extent, 3=to a great extent, and 4=to a very great extent).

[c]Rated on a 1-6 forced Likert scale (1=disagree strongly, 2=disagree moderately, 3=disagree slightly, 4=agree slightly, 5=agree moderately, and 6=agree strongly).

Qualitatively, students valued peer feedback for its ability to capture nonverbal cues and provide a patient-centered perspective. One student noted, “I think peer feedback considers the whole session, including tone, nonverbals, use of silence, etc, so it feels more relevant as a whole.”

### *RQ2: Design Limitations*

Students identified both technical and substantive limitations in CARE’s design. Transcription errors, often linked to audio quality, sometimes undermined the utility of feedback: “There are occasional issues with the transcript... CARE is unable to identify or differentiate between patient and patient at times, resulting in occasionally inaccurate feedback.” Beyond technical issues, participants described the scope of feedback as narrow, with repeated emphasis on empathy and validation while overlooking other therapeutic strategies. As one noted, “Most of the feedback was suggesting that I empathize with and validate the patient, but there are many other techniques... which the AI doesn’t seem to recognize.” Others emphasized that responses were evaluated in isolation: “It feels like each response is being considered separate and apart from the rest, so it feels less relevant to the content overall.”

Despite these concerns, students highlighted the value of alternative phrasing and immediate suggestions. One remarked, “I like the alternative responses... it gives me ideas for how to better and more concisely word my responses.” Another added, “I appreciate how content-oriented AI is! I find the alternative responses extremely helpful as I can use certain phrases in future sessions.” While some described the system as “robotic” or “less personal,” others appreciated its consistency, contrasting it with the subjectivity of human feedback: “Human feedback might only apply for that individual client... I think AI gives broader and more objective feedback that applies to the situation instead of the person.”

### *RQ3: Course Integration*

The integration of CARE into the curriculum was shaped by instructional stance and student perceptions of the research team. In the first course, CARE was introduced under exploratory conditions, which encouraged discussion. Use of CARE’s role-play transcript feedback was optional, and its role was described as exploratory. Of the 29 students enrolled in the first course, 14 (48%) completed surveys about the integration of CARE into the course, and 9 (31%) provided comparative ratings about peer roleplay versus CARE.

When asked what prevented them from using CARE during the course, students most frequently cited concerns about privacy and data security, a preference for receiving feedback from a human rather than AI, uncertainty or lack of guidance on how to use the tool, and apprehension about the nature of the research being conducted by the research team (eg, “making a chatbot therapist”).

### *RQ4: Acceptance*

A key concern raised by students in the first course was the security and intended use of their video transcripts when uploaded to the CARE platform. To address these concerns, the study team provided detailed explanations during class and in the frequently asked questions (FAQs) document (Multimedia Appendix 5). Students were informed that all video transcripts would be stored on a secure, password-protected cloud server. In rare cases where transcription errors occurred, a member of the research team would access the server to identify and troubleshoot the affected audio file. Aside from this limited quality control function, students were assured their audio files would not be used for any other purpose and would be deleted upon the completion of research analyses. Similarly, text transcripts were to be stored on a secure, password-protected server within CARE’s internal database. These transcripts were accessible only to the student and their selected peer partner. If students opted to use the “Review Feedback” tool, their transcripts were processed by CARE’s AI feedback model.



XSL•FO
RenderX

Importantly, CARE's model is entirely self-hosted and does not rely on commercial AI services (eg, ChatGPT), ensuring that student data are not used to train external proprietary models.

In addition to these measures, students were provided with a formal study information sheet outlining all data protections and ethical commitments associated with the project.

## Second Course: DP

### *Overview*

Building on the first course, the second course study shifted focus to the use of CARE's voice-based AI-simulated patients for DP, comparing this to Skillsetter's video-based practice. CARE enabled dynamic, back-and-forth dialogue with AI-simulated patients, whereas Skillsetter required a single, one-time response to a video prompt.

Students had already been introduced to the CARE platform in the prior term. The difference for the second course was that students could use CARE for DP, speaking with AI-simulated patients. Students could also receive AI feedback on the basic counseling skills used in their interactions with the AI-simulated patients. However, the focus was on the utility of the DP itself rather than the feedback.

### *RQ1: Comparative Value of CARE Voice-Based AI-Simulated Patients and Skillsetter Video Vignettes*

As presented in Table 5, the only significant difference between modalities was learning from mistakes, which students rated higher for Skillsetter than CARE (mean 4.52, SD 2.00 vs mean 3.20, SD 2.04; $P$=.03; $d$=0.49). Descriptively, the remaining comparisons diverged in direction rather than favoring one modality overall: Skillsetter was rated slightly higher for immersion (mean 4.16, SD 1.97 vs mean 3.88, SD 2.05) and confidence responding spontaneously (mean 4.52, SD 2.10 vs mean 4.32, SD 1.82), whereas CARE was rated slightly higher for communication beyond words (mean 4.20, SD 1.96 vs mean 3.80, SD 2.02) and building a therapeutic alliance (mean 4.52, SD 1.85 vs mean 3.80, SD 2.26). These differences were small relative to the variability in responses (SDs ≈ 1.8 - 2.3), and none approached significance. Qualitatively, students reported that CARE's AI-simulated patients provided a "good opportunity to practice using the skills learned in class" and allowed for interactive dialogue, which is not possible with Skillsetter. One student noted they used the AI-simulated patients because "you don't get to explore the back and forth with Skillsetter."

**Table .** Perceived value of challenging patient scenario practice with traditional video vignettes and instructor feedback (Skillsetter) compared to voice-based AI-simulated patients and AI feedback (CARE) rated by students in the second course (N=25).

| Question and modality | Mean (SD) | Median (IQR) | $P$ value[a] |
|---|---|---|---|
| Learning from mistakes[b] | | | .03 |
| CARE | 3.2 (2.041) | 3 (1-5) | |
| Skillsetter | 4.52 (2.002) | 5 (3-6) | |
| Immersion in interaction[b] | | | .66 |
| CARE | 3.88 (2.048) | 4 (2-5) | |
| Skillsetter | 4.16 (1.972) | 4 (2-6) | |
| Communication beyond words[b] | | | .58 |
| CARE | 4.2 (1.958) | 4 (3-6) | |
| Skillsetter | 3.8 (2.021) | 4 (2-5) | |
| Confidence responding spontaneously[b] | | | .80 |
| CARE | 4.32 (1.819) | 4 (3-6) | |
| Skillsetter | 4.52 (2.104) | 5 (2-7) | |
| Building a therapeutic alliance[b] | | | .13 |
| CARE | 4.52 (1.851) | 5 (3-6) | |
| Skillsetter | 3.8 (2.255) | 4 (2-6) | |

[a]$P$ values come from Wilcoxon signed-rank tests comparing CARE and Skillsetter modalities.

[b]Rated on a 1-7 scale (1=strongly disagree, 2=disagree, 3=somewhat disagree, 4=neither agree nor disagree, 5=somewhat agree, 6=agree, and 7=strongly agree).

While the second course did not include a peer role-play component, some students reflected on peer roleplay in the open-ended survey responses in course 2. CARE was also valued for simulating difficult scenarios that would be hard to practice with peers, such as dealing with an "angry client, silent client." Despite these benefits, a strong preference for practicing with humans remained. Students felt that AI practice could not replicate the nuances of human interaction, with one stating, "Nothing replaces role plays." There were also concerns that skills developed with the AI might not transfer to real therapy,



XSL•FO
RenderX

with one student stating they had "learned how to orient towards receiving positive feedback from CARE rather than toward providing thoughtful and nuanced support for a human person."

### *RQ2: Design Limitations*

The most prevalent technical concern (20/25, 80% of respondents) was that the system does not detect vocal pauses, indicating a significant gap in the speech recognition capabilities. More than half of respondents (14/25, 56%) felt that the AI-simulated patients did not feel realistic, with a similar number (13/25, 52%) noting limited emotional expression in the voices. Transcript inaccuracy (10/25, 40%), robotic-sounding voices (9/25, 36%), lack of visual avatar representation (7/25, 28%), and voice recognition issues (6/25, 24%) were additional functionality concerns raised by participants.

Students expressed consistent concerns about the quality and personalization of AI feedback. The 3 top concerns received the same high level of response (17/25, 68%): students wished the AI-simulated patient could provide feedback from its perspective on how the student, as a therapist, made them feel. Students desired a system that tracked and referenced their progression over time. Many reported that the AI was providing unhelpful feedback. Nearly two-thirds (16/25, 64%) of respondents wished the feedback addressed the session as a whole. There was a desire for greater variety in evaluation criteria (15/25, 60%) and feedback that incorporates vocal elements rather than just written transcripts (13/25, 52%).

In response to students' concerns about the limited utility of CARE's feedback, the teaching team implemented a supplemental feedback process in which the TA provided individualized written feedback on students' CARE transcripts. Each week, 3 to 5 students were randomly selected to receive this feedback, which was designed to enhance learning by integrating human insight with AI-generated suggestions. The TA provided weekly written feedback to each student who elected to use CARE, then emailed students individually to ask permission to post their feedback, identified by name, on the course Canvas (Instructure) page. Consenting students' feedback was posted to a discussion thread so that all students could view it, contribute to the discussion, and potentially benefit from it. The decision to supplement CARE in this way was informed by feedback indicating that the AI's responses alone were often perceived as insufficient for skill development. This feedback was obtained and documented in classroom memos by the TA.

One student reflected on the benefits of this hybrid model:

> *Additionally, I think everything involved with AI would have a better outcome when there is that human aspect. [TA] ended up providing additional feedback for us weekly for CARE. This combined the AI aspect with the human aspect, so it wasn't only AI that we were interacting with. I feel like this maximized our learning. It would also have been interesting if we actually looked at some of the AI feedback and analyzed those using a forum, in addition to human feedback. I think doing so would again emphasize the aspect of human and AI [or AI-assisted human feedback], instead of just AI doing all the work. AI's benefits wouldn't come alive if there weren't human input to modulate it.*

This perspective underscores students' desire for feedback that reflects both the analytical consistency of AI and the nuance of human clinical judgment.

### *RQ3: Factors About Course Integration That Impacted Acceptability*

The strongest concerns about CARE's integration into the classroom centered around how AI might replace human elements of education (Table 6), with 22 (88%) of 25 respondents worried about AI replacing peer support and feedback, and 20 (80%) of 25 respondents concerned about AI replacing peer interaction. Many students (19/25, 76%) questioned whether they were simply user-testing a product rather than engaging in an educational experience. Students raised concerns about potential environmental impact (16/25, 64%), suggesting awareness of the computational resources required for AI systems, including potential environmental impacts.

**Table .** Concerns about how CARE was integrated into the classroom experience.

| Concern | Choice count (N=25), n (%) |
|---|---|
| AI replacing peer support and feedback | 22 (88)[a] |
| AI replacing peer interaction | 20 (80)[a] |
| Am I just doing user testing for a product | 19 (76)[a] |
| Environmental impact | 16 (64)[a] |
| How my data is used by the research team | 14 (56)[a] |
| What if they are making an AI therapist to replace me | 12 (48) |
| I feel that my data is going to be used to make an AI therapist | 12 (48) |
| Other concerns | 7 (28) |

[a]Values indicate fields where more than half of the 25 respondents indicated concerns, with students able to select multiple concerns.

Data usage concerns were prominent, with 14 (56%) of 25 respondents worried about how their data are being used by the research team, and 12 (48%) of 25 respondents concerned about their data potentially being used to train AI therapists.



XSL•FO
RenderX

### *RQ4: Acceptance*

#### Overview

In addition to items about CARE, the second-course survey asked about AI in psychotherapy education and clinical practice generally. The responses below address this broader domain and should not be read as evaluations of CARE. Students articulated a range of perspectives, coalescing around three thematic domains: (1) openness to AI integration in training and administrative contexts, (2) boundary concerns regarding the application of AI in clinical care, and (3) broader ethical apprehensions related to AI's societal impact.

Several students expressed cautious optimism about the integration of AI into psychotherapy, particularly in training contexts and administrative functions. For these individuals, AI was not viewed as a threat to the therapeutic profession, but rather as a supplemental tool that could enhance learning and reduce burdens associated with nonclinical tasks. As one student noted, "I do not personally hold the fear of AI replacing therapists, so the training and administrative help resonates with me." Another student highlighted the relevance of technological adaptability in a rapidly evolving professional landscape:

> *As trainees in therapy, I truly see the importance of adapting to these changes and evolving with our society at large. Eventually, there will be other therapists who utilize AI, and we will be competing with those who know how to utilize it.*

#### Openness to AI Integration in Training and Administrative Contexts

Administrative utility was a recurring subtheme, with several students noting the promise of AI in alleviating logistical tasks such as billing, insurance coordination, and website maintenance. One student proposed that "a better focus for AI in psychotherapy would be having an AI handle the billing, insurance, outreach, and website management that is non-billable time for therapists, rather than training." Others concurred, suggesting that AI could support clinical infrastructure without compromising the therapeutic relationship.

#### Boundary Concerns Regarding the Application of AI in Clinical Care

Despite general receptivity to AI in administrative and educational settings, many students expressed strong reservations about potential clinical applications affecting direct patient care. A recurrent concern was that incorporating AI into training contexts might normalize its use in psychotherapy delivery, thereby facilitating a negative chain of future events. One student remarked, "Honestly, [I am] a bit more concerned, as I think while the benefits are there, it is a slippery slope." Another reflected, "I still feel uncertain about... the possibility of AI eclipsing in-person therapy in the future."

This apprehension was often anchored in the perceived irreplaceability of human relational dynamics. As one student articulated, "Nothing can replace in-person practice and human interaction. Beyond certain admin tasks, I do not think AI should be used in clinical practice." Another student emphasized the importance of ethical containment: "It's still important to be cautious with it, and it should never replace an actual human in any context other than training." These sentiments underscore a shared appreciation for the therapeutic alliance and a desire to establish clear boundaries for emerging technologies.

#### Broader Ethical Apprehensions Related to AI's Societal Impact

A third thematic domain reflected broader ethical considerations about AI's proliferation across societal and global systems. For some students, skepticism about AI in psychotherapy was rooted not solely in clinical concerns but also in a larger critique of AI's sociopolitical implications. One student remarked, "It has increased my concerns about the increase of AI use throughout the world." Another described their initial discomfort with the study's AI component, stating, "As someone who has bigger-picture ethical concerns about AI, in relation to environment, jobs, and broader social/societal impact, I was initially pretty upset about having CARE be a required part of this course."

Concerns about commercialization, technological determinism, and the disconnect between developers and end users were also prominent. One student stated, "The lack of awareness about therapy for those who are developing models frankly terrifies me," while another wrote, "I feel very hesitant about the use of AI in any space in psychotherapy. To me, it feels like too slippery of a slope, and from history, it seems like things like this inevitably end up in the wrong hands, even if that is not the starting intention." Reflecting these reservations, some students reported electing not to use the AI platform for training, citing "broader ethical concerns around AI and its rapid expansion."

Together, these themes reflect a complex and multidimensional stance among emerging psychotherapists. While students acknowledged the potential utility of AI in specific educational and administrative contexts, they also articulated clear boundaries for its use in clinical care and raised critical ethical questions about its integration into psychotherapy and society more broadly.

## Discussion

### Principal Findings

This study examined how doctoral psychotherapy students experienced AI-augmented DP embedded across 2 required courses in a first-year psychotherapy sequence at an APA-accredited clinical psychology program. The 2 courses involved different configurations. The first course served as a formative pilot in which students completed weekly peer role-plays, and CARE was introduced only as an optional source of AI-generated feedback on the resulting role-play transcripts, supplementing the feedback students already received from their peer partners. The second course introduced AI-simulated patients as a practice modality and placed 2 bundled packages side by side, matched for weekly curricular objectives: CARE, comprising multiturn spoken dialogue with voice-based LLM-driven AI-simulated patients followed by automated feedback, and Skillsetter, comprising single-turn recorded responses to standardized video vignettes performed by actors, with instructor and TA feedback.



XSL•FO
RenderX

Three findings stand out across the sequence. First, uptake appeared to be governed less by platform performance than by students' prior dispositions toward AI. In the first course pilot, 23 (79%) of 29 students uploaded at least one peer role-play transcript, but privacy concerns and a preference for human feedback were the reasons most frequently cited by those who did not, and comparable apprehension led the instructional team to adjust the second course's initial requirement after the first week, so that students could elect to complete additional Skillsetter practice sessions instead of CARE. Second, students' comparative ratings of the 2 second-course packages converged on most dimensions, with only learning from mistakes differing significantly in Skillsetter's favor, while their open-ended responses expressed a clear and consistent preference for practicing with humans. Third, the concerns students raised were largely not instructional; the most frequently endorsed were that AI would displace peer support and peer interaction. Taken together, these findings suggest that the principal obstacles to embedding AI-augmented practice in this curriculum were relational and ethical rather than technical, although technical limitations were substantial and may have compounded them.

### Insights From the First Course Pilot

Students rated peer feedback significantly above CARE's AI feedback on being constructive and helpful, and on recognizing areas to improve, attributing peers' advantage to the capacity to integrate tone, nonverbal information, and a patient-centered stance. As CARE operated on transcripts, much of what students valued in peer feedback was information the system could not access, a constraint shared by LLM-based feedback systems relying on textual representations of therapy [12,22,44]. The comparison may therefore have been less between human and machine judgment than between judgment informed by the full session and judgment informed by a degraded record of it.

Students' explanations nonetheless resemble 2 documented patterns. Research on algorithm aversion finds that people lose confidence in algorithms more rapidly than in people after comparable errors [45], which may help explain why transcription failures appeared to discredit CARE's feedback more broadly than those specific errors would warrant. Longoni et al [46] report that resistance to medical AI is associated with uniqueness neglect, the concern that automated systems cannot accommodate individual circumstances; our students described AI feedback as evaluating responses "separate and apart from the rest," while peer feedback "considers the whole session." We draw on these literatures interpretively because our design tested neither mechanism. Notably, 1 student inverted the framing entirely, valuing AI for giving "broader and more objective feedback that applies to the situation instead of the person," a divergence cautioning against treating preference for human feedback as uniform.

### Comparative Value of AI-Simulated Patients and Video Vignettes

Students' ratings of the 2 packages appeared to converge on most dimensions. Skillsetter was rated significantly higher only for learning from mistakes, while immersion, confidence responding spontaneously, communication beyond words, and building a therapeutic alliance did not differ significantly, the latter 2 showing numerically higher means for CARE. Given wide variability and limited power, these patterns are best read as descriptive.

Students also reported concerns about realism, with 56% (14/25) indicating that the AI-simulated patients did not feel realistic, and 52% noting limited vocal expression. Simulation scholarship suggests one possible reading, having generally found fidelity to be largely independent of educational effectiveness, and proposed that functional task alignment may explain outcomes better than physical resemblance [47], with work in mental health simulation reporting psychological fidelity as more critical to learning than physical fidelity [48]. Students may perhaps have experienced the simulated patients as unconvincing depictions of people, while still finding the task aligned with what the course asked them to rehearse.

A notable feature of the open-ended responses was that students introduced a comparison the survey did not ask about. The second-course items concerned CARE and Skillsetter only, and peer role-play was not a component of that course, yet several students referenced it unprompted, with 1 respondent stating that "nothing replaces role plays." One interpretation is that live practice with a person functioned as an implicit reference point against which both packages were evaluated, although neither offered it. This would be broadly consistent with reports from other relational health professions, where trainees describe AI as acceptable in a supportive role but maintain that it cannot substitute for human-centered interaction and therapeutic communication [49,50]. Some students additionally expressed concern that practicing with AI might lead them to optimize for system-specific feedback rather than responsiveness to a person, an apprehension that appears to parallel deskilling concerns noted in health professions education [51]. Ratings and written responses thus seemed to diverge, with the latter suggesting a preference that the comparative items may not have been positioned to capture.

### Identity-Based Rather Than Integrity-Based Resistance

The concerns our students raised differ in kind from those documented among students in general higher education, where reservations about generative AI cluster around accuracy, academic integrity, and overreliance diminishing critical thinking [52]. These were largely absent from our data. No student framed CARE as a shortcut around effort or worried that using it would undermine their own learning. Their most endorsed concerns were that AI would replace peer support and feedback (22/25, 88%) and peer interaction (20/25, 80%), and nearly half worried their data might build an AI therapist that could replace them. The object of concern appears to have been not their academic conduct but their professional future.

This pattern resembles reports from other relational health professions more than from general education. A systematic review of nursing students and practicing nurses found consistent argument that AI should augment rather than replace human empathy and judgment, with fear of deskilling impeding adoption [49]; qualitative studies describe AI accepted as supportive but incapable of replacing therapeutic communication [50]. Jussupow et al [53] offer a possible mechanism, reporting



XSL•FO
RenderX

that threats to professional recognition and to professional capabilities predicted resistance among medical students and physicians, with students experiencing stronger identity threats than experienced clinicians. Students whose developing expertise is defined by relational work may appraise AI in that domain partly as a threat to the identity they are acquiring rather than solely as a tool of uncertain utility. If so, this resistance may call for different responses than those developed for general education. Guidance on academic integrity and appropriate reliance addresses concerns our students did not appear to hold; what they voiced was uncertainty about whether the relational core of their profession would be preserved. Such concerns seem more tractable through transparency about a system's purposes, framing that positions AI as supplementing rather than substituting for human relationship, and protection of the peer and supervisory interactions students valued most. We note the limits of this comparison: we sampled 1 cohort in 1 program and rest the contrast on triangulation against published work rather than on data of our own.

### Design Limitations and Opportunities

Students' design concerns appeared to center on feedback that did not attend to the session as a whole. Three were endorsed equally often (17/25, 68% each): that the simulated patient could not report how the student's responses made it feel, that progress was not tracked over time, and that feedback was frequently unhelpful; 64% (n=16) wanted feedback addressing the whole session. Prior work suggests that pairing simulated practice with feedback is both desired [23,26] and associated with skill improvement [21,22]; sustained multiweek use may additionally surface a need for criteria matched to each week's pedagogical goal rather than to a fixed set of listening microskills.

Across both courses, students seemed to prefer not better AI feedback but AI feedback accompanied by human feedback. Peer feedback was rated above AI feedback in the first course, and supplemental TA feedback on CARE transcripts appeared to be received positively in the second, with 1 student observing that "AI's benefits wouldn't come alive if there weren't human input to modulate it." Hybrid approaches have been examined in general and online education [54,55] and in mental health support [17-19]. As we did not measure skill, our findings cannot speak to whether hybrid feedback improves acquisition, but they may suggest that automated feedback alone was not readily accepted as a substitute in this setting.

### Course Integration

Instructional stance played a critical role in shaping students' engagement. Concerns about data security and transcript use were salient, leading the team to develop a detailed FAQ and formal information sheet to clarify protections. These lessons guided adjustments in the second course, where CARE was more fully embedded as a core pedagogical component. Incorporating CARE as a weekly course requirement corresponded to higher participation and normalization of its use. However, after the second week of the second course, CARE became optional due to students' apprehension. They could choose to use CARE for DP or complete additional Skillsetter videos.

This phenomenon aligns with other recent course-based integrations of AI-simulated patients that vary in instructional requirements. Prior work has documented both optional and required approaches to AI-supported practice, with optional implementations associated with more limited uptake and required implementations designed to ensure broad exposure across learners [25,26]. Taken together, these approaches suggest that optional integration may limit reach, whereas required engagement may increase participation and normalize use. At the same time, the present findings indicate that when AI-based practice is required, sustained engagement may depend on parallel instructional investment, including responsiveness to student concerns and the provision of human feedback to contextualize and support AI-augmented learning.

### Acceptance and Broader Perspectives

In addressing acceptance of AI in psychotherapy education (RQ4), students articulated views that were neither uniformly negative nor organized around instructional considerations alone. Several expressed cautious openness, particularly toward training and administrative applications, with 1 proposing that AI might more appropriately handle billing and insurance coordination than training itself. This ambivalence appears broadly consistent with reports that health professions students often hold simultaneously positive and apprehensive attitudes toward AI [56], and with findings in nursing that trainees accept AI as a supportive tool while resisting it in more central roles [49,50]. It also cautions against characterizing this cohort as uniformly resistant.

Where students did express reservations, these seemed to concentrate at the boundary between training and clinical care rather than within training itself. A recurrent concern was that normalizing AI in coursework might facilitate its later migration into practice, described by 1 student as "a slippery slope," and such apprehension was typically anchored in the perceived irreplaceability of human relational processes widely regarded as foundational to psychotherapy [1,4,35]. Notably, students drew this line at clinical care rather than at learning: their objection appeared to be less that AI cannot support skill development than that its presence in training might legitimize its presence elsewhere. This may distinguish their concerns from those documented in general higher education, where reservations more often center on integrity and overreliance affecting students' own learning [52].

A third set of concerns extended well beyond psychotherapy and was not anticipated in our design. Students raised commercialization and the perceived disconnect between developers and practitioners, with one writing that "the lack of awareness about therapy for those who are developing models frankly terrifies me." Environmental impact was endorsed by 64% (n=16) of respondents, a theme that appears to have received limited attention in the psychotherapy training literature to date, and some students reported declining to use the platform partly on these grounds. A few spontaneously referenced the January 2025 Southern California wildfires; because this association was not systematically assessed, we note it only as possible context. Taken together, these responses may suggest that acceptance of a particular educational tool is partly



XSL•FO
RenderX

conditioned by students' appraisal of the broader industry producing it, a consideration that instructional design alone may be poorly positioned to address.

## Limitations

Several limitations should be noted. This study was conducted within a single graduate training program, and the experiences of this cohort may reflect the particular pedagogical environment, instructional style, or program culture in which CARE was introduced. The program's highly selective nature, which admitted 30 students from 644 applicants during the 2024 to 2025 cycle (a 7.1% admit rate), may further limit generalizability to less clinically experienced graduate students. In addition, the sample consisted of first-year graduate students enrolled in a multiyear program characterized by intensive supervision, including regular review of recorded therapy sessions with experienced supervisors. This context likely shaped student perceptions of AI-assisted training and may limit generalizability to settings where trainees have fewer opportunities for expert feedback (eg, peer counselors or paraprofessionals with limited supervision access).

Beyond the characteristics of the program itself, participation varied within the cohort. The first-course survey was optional and completed by 14 (48%) of 29 students; the second-course survey was a required component of the coursework and completed by 25 (86%) of 29. Students who used CARE and completed the first-course survey may differ systematically from those who did not, particularly given the privacy and data security concerns reported by students who declined the tool, so the 9 students providing comparative ratings of AI and peer feedback are not representative of the cohort. The first course was also shorter, spanning 6 weeks rather than 10, and CARE was an optional supplement rather than an assigned activity. We report these findings as preliminary and include them to provide context for the second-course deployment. Throughout, we describe findings as those of survey respondents rather than of the cohort.

The CARE platform integrated existing technologies for voice-based AI-simulated patients and feedback generation that were best performing in December 2024. Technical issues may have reflected the specific version deployed rather than inherent limitations of AI-assisted training tools: voice-based practice interactions had issues with transcription accuracy, speech recognition, and limited emotional expression, while feedback was limited to skill-based content on patient-centered listening skills rather than scenario-specific guidance regarding therapeutic-alliance challenges.

Our safety procedures relied on the courses' existing supervisory structures rather than on study-specific monitoring. This design has a defensible rationale: the topics were already part of the curriculum, taught by the same faculty, and practiced concurrently in an established platform, but it means we cannot report a systematic account of whether the AI-simulated patients generated inappropriate content or whether students experienced distress that went unvoiced. Deployments at larger scale, or in settings without an embedded clinical instructor, would require monitoring procedures we did not implement here: routine transcript review, automated flagging of high-risk content, or structured post-session check-ins.

A central constraint on interpretation is that CARE and Skillsetter differed simultaneously along several dimensions rather than a single manipulated variable. The 2 packages contrasted voice-based with video-based interaction, multiturn dialogue with single-turn response, and a research prototype with an established commercial platform. Because these attributes were bundled, the present design cannot isolate the educational contribution of any one of them, including AI per se, voice interaction, dialogic format, feedback source, or platform maturity. We therefore frame our findings as an evaluation of students' perceptions of 2 bundled instructional packages as actually delivered in the classroom, rather than as a test of "AI versus traditional training." The comparative ratings should be read as package-level judgments; disentangling the individual active ingredients will require factorial or component-controlled designs in future work.

All outcomes reported here are student self-report. We did not assess objective change in psychotherapy competence, skill acquisition, alliance-building ability, or clinical readiness, and we make no claims that CARE improved these. Perception items (eg, feeling "more equipped to build a therapeutic alliance") index students' subjective experience, not measured skill.

## Future Work

The findings and limitations of this study suggest several priorities for future research on AI-assisted psychotherapy training. Short-term work should focus on improving the technical functionality of AI platforms. Students in this study consistently identified transcription errors as well as AI-simulated patients with limited vocal expressiveness as barriers to engagement. Research is needed to evaluate whether improvements in speech recognition, naturalistic prosody, and affective responsiveness increase the realism and pedagogical value of AI-simulated patients.

Parallel efforts should advance feedback systems toward greater context awareness. Current utterance-level approaches offer timely guidance on specific communication skills, but models that process turns independently can produce feedback that is repetitive, contradictory, or disconnected from the broader therapeutic process. Prior work on automated feedback and performance assessment in psychotherapy and counseling has similarly noted that utterance-level or narrowly scoped feedback may be insufficient for supporting integration of relational and process-oriented skills across the flow of a session [12,19,20]. Approaches that consider the entire session history and enforce global consistency during feedback generation could better support students in developing coherent, session-oriented improvements. Moreover, consistent with emerging evidence that human–AI collaborative feedback may enhance contextual understanding and learner trust [17-19], short-term research should further evaluate hybrid feedback models that combine AI-generated suggestions with human commentary, an approach that students in the present study viewed as particularly beneficial.



XSL•FO
RenderX

Beyond technical functionality, future studies should evaluate whether AI-assisted training translates into measurable improvements in clinical skill acquisition. Recent research on AI-supported counseling and psychotherapy training has highlighted the need to move beyond learner self-report toward objective indicators of competence, including supervisor ratings, behavioral coding of microskills, and structured performance assessments [6,21,22]. Critically, such assessments should extend beyond generic skill-based evaluations (eg, patient-centered listening) to include scenario-specific rubrics aligned with the particular competencies being taught, whether therapeutic-alliance navigation, crisis intervention, or other course-specific learning objectives. Experimental designs that directly compare AI-only, human-only, and hybrid feedback approaches would clarify the specific pedagogical contributions of each modality and help determine whether AI-supported practice confers additive benefits beyond established training methods. Multisite studies are also essential to assess whether findings generalize across programs with differing curricula, instructional philosophies, and trainee populations, a limitation frequently noted in early evaluations of AI-based training tools [6,21].

At a broader level, long-term research should address how AI integration reshapes the pedagogy and culture of psychotherapy training. Questions include whether repeated practice with AI-simulated patients influences professional identity formation and sensitivity to relational nuance. Ethical considerations will remain central, particularly regarding transparency in data use, the potential displacement of human elements in training, and environmental sustainability. The salience of ecological concerns in this study, possibly amplified by the January 2025 Southern California wildfires, highlights the need for future work to consider how external sociocultural contexts intersect with classroom-based experiences of AI. Ultimately, research must examine whether AI can be a useful adjunct to psychotherapy training, enhancing current methods, or whether its widespread adoption risks altering the relational foundations of psychotherapy education.

## Recommendations for Classroom Implementation

These recommendations were informed by student feedback across both terms, as well as by the adjustments made between the first course pilot and the second course deployment.

### *Transparency*

Clear, proactive communication about data use, research protocols, and pedagogical aims is essential to building student trust. The second course implementation benefited from refinements to communication strategies developed in response to concerns raised during the first course pilot.

### *Hybrid Models*

Combining AI feedback with human oversight may mitigate concerns about impersonality and enhance learning outcomes. The addition of supplemental TA feedback during the second course exemplified this adjustment and was positively received by students. Future research may benefit from systematically examining the impact of required AI integration on students' perceptions of the teaching team, including measures of instructional alliance, trust, and perceived support, to assess whether and how AI-mediated feedback influences the relational dynamics of the learning environment.

### *Instructor Stance*

Student reflections suggested that instructional framing influenced their engagement with CARE. In the first course, the tool was positioned as optional following student concerns and apprehension, and several students reported feeling more willing to try it under these conditions than when it was framed as required. In the second course, CARE was initially presented as a central, required component of the course, which corresponded with higher participation early in the term. However, after concerns were raised, the instructional team shifted to making it optional, where students could engage in additional Skillsetter practice instead of CARE. These observations suggest that how the AI tool is introduced and framed by instructors plays a role in shaping student engagement, and that providing choice in an assigned activity can foster greater willingness to engage with AI-assisted training.

## Conclusions

The integration of CARE across 2 academic terms illustrates both the promise and challenges of AI-assisted psychotherapy training. The first course pilot surfaced key concerns regarding feedback quality, course integration, and data transparency, while the second course study demonstrated how iterative adjustments, such as clearer communication, hybrid feedback models, and shifts in instructional framing, enabled more effective implementation. At the same time, students expressed persistent preferences for human-based training and raised concerns about realism, data use, and broader societal implications, including environmental impact. Taken together, these findings suggest that at this stage of development, AI has the potential to serve as a valuable adjunct to psychotherapy education when framed as a supplement to human interaction, implemented transparently, and paired with human oversight.

Looking ahead, priorities for advancing this field can be organized across 3 levels. Short-term priorities involve addressing technical limitations and refining feedback mechanisms. Medium-term priorities focus on evaluating whether AI-assisted practice translates into measurable improvements in clinical performance, ideally through controlled, multisite studies that include objective learning outcomes. Long-term priorities include examining how AI integration reshapes the pedagogy and culture of psychotherapy training, professional identity formation, and ethical considerations such as data use, role displacement, and environmental sustainability.

The utility of AI-augmented psychotherapy training may differ across educational contexts. In this graduate program, students receive intensive supervision, including live or taped review of clinical encounters, which may shape how they perceive the added value of AI. In other settings, such as programs with more limited supervisory resources, AI tools may assume a more prominent role in supporting DP and providing structured feedback.



XSL•FO
RenderX

Thus, the pedagogical value of AI may not be uniform but instead contingent on the broader training environment, with its potential relevance varying according to the availability of human supervision and feedback.

By situating AI within authentic classroom contexts and iteratively adapting implementation based on student feedback, this study provides a strong foundation for future inquiry in psychotherapy pedagogy. Its contribution lies not only in documenting students' perceptions of AI-assisted training but also in demonstrating how responsive piloting can inform best practices for responsibly integrating emerging technologies into psychotherapy education.

## Acknowledgments

We thank the course instructors, teaching assistants, and students for their collaboration. We especially thank Alan Zhang and Raj Sanjay Shah for their early contributions during summer 2024 for helping to develop and architect the software infrastructure that supported CARE's continuous, multiweek deployment.

## Generative AI Use

Generative AI tools were used by the authors when writing data analysis code, developing prompts, and editing the manuscript.

## Funding

This study was funded by the Stanford Psychiatry and Behavioral Sciences Department Innovator Grant, Stanford HAI Seed Grant, and the Stanford Impact Labs Stage 1 Award. Funders did not influence the study design, analysis, interpretation, and decision to publish this research.

## Data Availability

The datasets generated and analyzed during this study, including survey and log data, are not publicly available because of the data sharing agreement with student participants; only aggregate or statistical analyses of course data are shared in scientific publications, including this manuscript. Data will be retained for 3 years (April 2025 to April 2028) to support scientific analyses of this deployment study by the authors, after which they will be deleted.

## Authors' Contributions

RL contributed to project conceptualization, obtaining funding, forming and maintaining cross-department collaboration, submitting the institutional review board (IRB), study planning, visiting the course to deliver information and answer questions about the study, conducting data analyses, and drafting and editing the manuscript. EC contributed to maintaining cross-department collaboration, conceptualizing integration of CARE into course curriculum, pilot-testing the CARE system, study planning, serving as a course TA (visiting the course to deliver information and answer questions about the study; providing TA feedback on a subset of assignments in winter 2025), conducting data analyses, and drafting and editing the manuscript. DS was the instructor of the fall course. She helped with cross-department team formation, provided feedback on course integration and survey design for the first course, and helped edit the entire article manuscript. JPP assisted with prompt design and pilot testing of the CARE's AI-simulated patients, as well as editing of the manuscript. WF assisted with prompt design and pilot testing of CARE's AI-simulated patients, as well as editing of the manuscript. JK was the instructor for the winter 2025 course. She provided feedback on course integration, survey design, and the manuscript. BA contributed to project conceptualization, obtaining funding, forming the cross-department collaboration, providing feedback on survey design, and editing the manuscript. DY contributed to project conceptualization, obtaining funding, forming the cross-department collaboration, and providing feedback on the IRB, study plan, and manuscript.

## Conflicts of Interest

Three of the authors (RL, BA, and DY) are part of a team that developed the CARE system and its components as detailed in prior publications. These authors are co–principal investigators or collaborators on internally and externally funded grants to study and adapt the CARE system to counseling training contexts. These authors do not have plans to file for intellectual property, commercialize the invention, or directly financially benefit from the CARE project. The insights from testing and deploying CARE inform the future academic plans for refining, adapting, and deploying CARE to new counseling education and training contexts. While CARE's underlying datasets and model artifacts have been publicly released in prior publications, the fully integrated CARE system remains closed source. Sharing of the fully integrated CARE software artifact is governed by a researcher collaboration agreement. This agreement dictates that future publications that build upon and adapt CARE's integrated system will (1) include these original authors, (2) not release the fully integrated system, but (3) can plan to publicly release component datasets or models that can benefit the research community as decided by the consensus of the collaborators.

Multimedia Appendix 1



XSL•FO
RenderX

Descriptions of the 12 AI-simulated patients used on the CARE platform in the second course, by week and topic, with each patient's vignette and target therapeutic skills.
[PDF File, 65 KB - i-jmr_v15i1e93199_app1.pdf ]

Multimedia Appendix 2
First-course midcourse survey: peer feedback and AI feedback. Questionnaire administered midway through the first course. It assessed CARE use and barriers to use, the perceived helpfulness of and comfort with peer and AI feedback, learning from reviewing a partner's AI feedback, and the influence of AI feedback on peer feedback exchanges. Items used Likert-type and open-ended response formats, and branching logic routed students according to whether they had received AI feedback on CARE.
[PDF File, 243 KB - i-jmr_v15i1e93199_app2.pdf ]

Multimedia Appendix 3
First-course end-of-course survey: peer feedback and AI feedback. Questionnaire administered at the end of the first course. It covers the same domains as the midcourse survey, with items addressing feedback from each of 2 peer partners and ratings of factors that make feedback exchange effective.
[PDF File, 370 KB - i-jmr_v15i1e93199_app3.pdf ]

Multimedia Appendix 4
Second-course end-of-course survey: comparing CARE and Skillsetter. Questionnaire administered at the end of the second course. It assessed general attitudes toward AI, concerns about CARE's AI patients, feedback, and classroom integration, and paired comparisons of CARE and Skillsetter across learning from mistakes, immersion, communication beyond word choice, spontaneous responding, and alliance building. It also asked about perceptions of AI feedback and experiences with utterance-level self-reflection. Items used 7-point Likert scales, multiple-choice, and open-ended responses.
[PDF File, 464 KB - i-jmr_v15i1e93199_app4.pdf ]

Multimedia Appendix 5
Frequently asked questions (FAQ) document on CARE data use and privacy. Information document provided to students in the first course addressing why no consent form was required, voluntary participation, how to raise questions, audio transcription, transcript storage and data use, how to request exclusion of specific transcripts, and the skills framework underlying CARE's AI feedback.
[PDF File, 1471 KB - i-jmr_v15i1e93199_app5.pdf ]

## References


1. American Psychological Association. Recognition of psychotherapy effectiveness. Psychotherapy 2013;50(1):102-109. [doi: 10.1037/a0030276]
2. Shedler J. Where is the evidence for "evidence-based" therapy? Psychiatr Clin North Am 2018 Jun;41(2):319-329. [doi: 10.1016/j.psc.2018.02.001] [Medline: 29739529]
3. Knox S, Hill CE. Training and Supervision in Psychotherapy: What We Know and Where We Need to Go: American Psychological Association; 2021. URL: https://psycnet.apa.org/record/2021-81510-010 [accessed 2026-09-25]
4. Wampold BE. How important are the common factors in psychotherapy? An update. World Psychiatry 2015 Oct;14(3):270-277. [doi: 10.1002/wps.20238] [Medline: 26407772]
5. Hill CE, Knox S. Psychotherapy Training and Supervision with Undergraduate and Graduate Students: American Psychological Association; 2023. URL: https://www.apa.org/pubs/books [accessed 2026-09-25]
6. Nurse K, O'Shea M, Ling M, Castle N, Sheen J. The influence of deliberate practice on skill performance in therapeutic practice: a systematic review of early studies. Psychother Res 2025 Mar;35(3):353-367. [doi: 10.1080/10503307.2024.2308159] [Medline: 38295223]
7. Larsson J, Werthén D, Carlsson J, et al. Does deliberate practice surpass didactic training in learning empathy skills? – A randomized controlled study. Nord Psychol 2025 Jan 2;77(1):39-52. [doi: 10.1080/19012276.2023.2247572]
8. Chen S, Wu M, Zhu KQ, Lan K, Zhang Z, Cui L. LLM-empowered chatbots for psychiatrist and patient simulation: application and evaluation. arXiv. Preprint posted online on May 23, 2023 URL: https://arxiv.org/abs/2305.13614 [accessed 2026-09-25] [doi: 10.48550/arXiv.2305.13614]
9. How it works. Skillsetter. 2024. URL: https://www.skillsetter.com/how-it-works [accessed 2026-09-25]
10. Louie R, Nandi A, Fang W, Chang C, Brunskill E, Yang D. Roleplay-doh: enabling domain-experts to create LLM-simulated patients via eliciting and adhering to principles. arXiv. Preprint posted online on Jul 1, 2024 URL: https://arxiv.org/abs/2407.00870 [accessed 2026-09-25] [doi: 10.48550/arXiv.2407.00870]
11. Wang R, Milani S, Chiu JC, et al. PATIENT- : using large language models to simulate patients for training mental health professionals. arXiv. Preprint posted online on May 30, 2024. [doi: 10.48550/arXiv.2405.19660]




XSL•FO
RenderX

12. Flemotomos N, Martinez VR, Chen Z, Creed TA, Atkins DC, Narayanan S. Automated quality assessment of cognitive behavioral therapy sessions through highly contextualized language representations. PLoS One 2021;16(10):e0258639. [doi: 10.1371/journal.pone.0258639] [Medline: 34679105]
13. Shah RS, Holt F, Hayati SA, et al. Modeling motivational interviewing strategies on an online peer-to-peer counseling platform. Proc ACM Hum-Comput Interact 2022 Nov 7;6(CSCW2):1-24. [doi: 10.1145/3555640]
14. Sharma A, Miner AS, Atkins DC, Althoff T. A computational approach to understanding empathy expressed in text-based mental health support. Presented at: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP); Nov 16-20, 2020. [doi: 10.18653/v1/2020.emnlp-main.425]
15. Xiao B, Imel ZE, Georgiou PG, Atkins DC, Narayanan SS. "Rate my therapist": automated detection of empathy in drug and alcohol counseling via speech and language processing. PLoS One 2015;10(12):e0143055. [doi: 10.1371/journal.pone.0143055]
16. Lin IW, Sharma A, Rytting CM, Miner AS, Suh J, Althoff T. IMBUE: improving interpersonal effectiveness through simulation and just-in-time feedback with human-language model interaction. arXiv. Preprint posted online on Feb 19, 2024 URL: https://arxiv.org/abs/2402.12556 [accessed 2026-09-25] [doi: 10.48550/arXiv.2402.12556]
17. Hsu SL, Shah RS, Senthil P, et al. Helping the helper: supporting peer counselors via AI-empowered practice and feedback. arXiv. Preprint posted online on May 15, 2023 URL: https://arxiv.org/abs/2305.08982 [accessed 2026-09-25] [doi: 10.48550/arXiv.2305.08982]
18. Sharma A, Lin IW, Miner AS, Atkins DC, Althoff T. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. Nat Mach Intell 2023;5(1):46-57. [doi: 10.1038/s42256-022-00593-2]
19. Chaszczewicz A, Shah RS, Louie R, Arnow BA, Kraut R, Yang D. Multi-level feedback generation with large language models for empowering novice peer counselors. Presented at: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Aug 11-16, 2024. [doi: 10.18653/v1/2024.acl-long.227]
20. Rudolph E, Seer H, Mothes C, Albrecht J. Automated feedback generation in an intelligent tutoring system for counselor education. Presented at: 19th Conference on Computer Science and Intelligence Systems (FedCSIS). [doi: 10.15439/2024F1649]
21. Louie R, Shah RS, Pacheco JP, Brunskill E, Yang D. Can LLM-simulated practice and feedback upskill human counselors? A randomized study with 90+ novice counselors. Presented at: CHI 2026: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems; Apr 13-17, 2026; Barcelona, Spain p. 1-31. [doi: 10.1145/3772318.3791821]
22. Tanana MJ, Soma CS, Srikumar V, Atkins DC, Imel ZE. Development and evaluation of ClientBot: patient-like conversational agent to train basic counseling skills. J Med Internet Res 2019 Jul 15;21(7):e12529. [doi: 10.2196/12529] [Medline: 31309929]
23. Steenstra I, Nouraei F, Bickmore T. Scaffolding empathy: training counselors with simulated patients and utterance-level performance visualizations. Presented at: CHI 2025: CHI Conference on Human Factors in Computing Systems; Apr 26 to May 1, 2025; Yokohama Japan p. 1-22. [doi: 10.1145/3706598.3714014]
24. Sim KYH, Lee RKW, Choo KTW. "Is this really a human peer supporter?": misalignments between peer supporters and experts in LLM-supported interactions. Proc ACM Hum-Comput Interact 2026 Oct 31;10(6):1-52. [doi: 10.1145/3816956]
25. Beeson ET, Zhai Y, Fulmer R, Burck AM, Maurya R. A pilot study evaluating the fidelity of ChatGPT in client simulations. JCPS 2025;19(3):1-16. [doi: 10.70013/d2a58d79]
26. Thesen T, O'Brien WN, Stone S, Pinto-Powell R. Generative AI as the first patient: practice, feedback, and confidence. Med Sci Educ 2025 Oct;35(5):2555-2560. [doi: 10.1007/s40670-025-02473-x] [Medline: 41555902]
27. Deliberate practice turns knowledge into skill. Skillsetter. URL: https://www.skillsetter.com/ [accessed 2026-09-25]
28. Hill CE. Helping Skills: Facilitating Exploration, Insight, and Action, 5th edition: APA; 2020. URL: https://www.apa.org/pubs/books/Helping-Skills-5e-Hill-Sample.pdf [accessed 2026-09-25]
29. Miller WR, Rollnick S. Motivational Interviewing: Helping People Change, 3rd edition: Guilford Press; 2013. URL: https://www.guilford.com/books/Motivational-Interviewing/Miller-Rollnick/9781609182274 [accessed 2026-09-25]
30. Norcross JC. Conclusions and recommendations of the interdivisional (APA divisions 12 & 29) task force on evidence-based therapy relationships. Society for the Advancement of Psychotherapy. 2014. URL: https://societyforpsychotherapy.org/evidence-based-therapy-relationships/ [accessed 2026-09-25]
31. Hartl TL, Zeiss RA, Marino CM, Zeiss AM, Regev LG, Leontis C. Clients' sexually inappropriate behaviors directed toward clinicians: Conceptualization and management. Professional Psychology: Research and Practice 2007;38(6):674-681. [doi: 10.1037/0735-7028.38.6.674]
32. Min DJ, Pérez-Rosas V, Resnicow K, Mihalcea R. PAIR: prompt-aware margin ranking for counselor reflection scoring in motivational interviewing. Presented at: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing; Dec 7-11, 2022. [doi: 10.18653/v1/2022.emnlp-main.11]
33. Pérez-Rosas V, Resnicow K, Mihalcea R, et al. VERVE: Template-based Reflective Rewriting.: EMNLP Findings; 2023. URL: https://aclanthology.org/2023.findings-emnlp.689/ [accessed 2026-11-05]
34. Shen S, Perez-Rosas V, Welch C, Poria S, Mihalcea R. Knowledge enhanced reflection generation for counseling dialogues. : ACL Presented at: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume




XSL•FO
RenderX

1); May 22-27, 2022; Dublin, Ireland p. 3096-3107 URL: https://aclanthology.org/2022.acl-long [accessed 2026-09-25] [doi: 10.18653/v1/2022.acl-long.221]
35. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychotherapy: Theory, Research & Practice 1979;16(3):252-260. [doi: 10.1037/h0085885]
36. Growing public concern about the role of artificial intelligence in daily life. Pew Research Center. 2023. URL: https://www.pewresearch.org/short-reads/2023/08/28/growing-public-concern-about-the-role-of-artificial-intelligence-in-daily-life/ [accessed 2026-09-25]
37. AMA physicians enthusiastic but cautious about health care AI: 2023 AMA augmented intelligence research survey. American Medical Association. 2023. URL: https://www.ama-assn.org/press-center/ama-press-releases/ama-physicians-enthusiastic-cautious-about-health-care-ai [accessed 2026-09-25]
38. García-Pérez MA. Use and misuse of corrections for multiple testing. Methods in Psychology 2023 Nov;8:100120. [doi: 10.1016/j.metip.2023.100120]
39. Hoffmann S, Lemster S, Collins G, et al. When to adjust for multiple testing: a unifying guiding principle. Biom J 2026 Aug;68(4):e70148. [doi: 10.1002/bimj.70148] [Medline: 42388011]
40. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
41. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health 2019 Aug 8;11(4):589-597. [doi: 10.1080/2159676X.2019.1628806]
42. Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qual Res Psychol 2021 Jul 3;18(3):328-352. [doi: 10.1080/14780887.2020.1769238]
43. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156. [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
44. García-Torres D, Fernández C, Mira JJ, Morales A, Vicente MA. Using AI-based virtual simulated patients for training in psychopathological interviewing: cross-sectional observational study. JMIR Med Educ 2025 Dec 23;11:e78857. [doi: 10.2196/78857] [Medline: 41433050]
45. Dietvorst BJ, Simmons JP, Massey C. Algorithm aversion: people erroneously avoid algorithms after seeing them err. J Exp Psychol Gen 2015 Feb;144(1):114-126. [doi: 10.1037/xge0000033] [Medline: 25401381]
46. Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. J Consum Res 2019 Dec 1;46(4):629-650. [doi: 10.1093/jcr/ucz013]
47. Hamstra SJ, Brydges R, Hatala R, Zendejas B, Cook DA. Reconsidering fidelity in simulation-based training. Acad Med 2014 Mar;89(3):387-392. [doi: 10.1097/ACM.0000000000000130] [Medline: 24448038]
48. Naismith LM, Kowalski C, Soklaridis S, Kelly A, Walsh CM. Participant perspectives on the contributions of physical, psychological, and sociological fidelity to learning in interprofessional mental health simulation. Simul Healthc 2020 Jun;15(3):141-146. [doi: 10.1097/SIH.0000000000000425] [Medline: 32218087]
49. El Arab RA, Alshakihs AH, Alabdulwahab SH, et al. Artificial intelligence in nursing: a systematic review of attitudes, literacy, readiness, and adoption intentions among nursing students and practicing nurses. Front Digit Health 2025;7:1666005. [doi: 10.3389/fdgth.2025.1666005] [Medline: 41079691]
50. Dagcan Sahin N, Yildirim M. Nursing students' experiences with artificial intelligence: a qualitative study on education, clinical practice, and future expectations. J Eval Clin Pract 2026 Mar;32(2):e70413. [doi: 10.1111/jep.70413] [Medline: 41833057]
51. Heudel PE, Crochet H, Filori Q, Bachelot T, Blay JY. Artificial intelligence in medicine: a scoping review of the risk of deskilling and loss of expertise among physicians. ESMO Real World Data Digit Oncol 2026 Jun;12:100693. [doi: 10.1016/j.esmorw.2026.100693] [Medline: 41890350]
52. Chan CKY, Hu W. Students' voices on generative AI: perceptions, benefits, and challenges in higher education. Int J Educ Technol High Educ 2023;20(1):43. [doi: 10.1186/s41239-023-00411-8]
53. Jussupow E, Spohrer K, Heinzl A. Identity threats as a reason for resistance to artificial intelligence: survey study with medical students and professionals. JMIR Form Res 2022 Mar 23;6(3):e28750. [doi: 10.2196/28750] [Medline: 35319465]
54. Kim J. Learnersourcing: improving learning with collective learner activity. : Massachusetts Institute of Technology; 2015 URL: http://hdl.handle.net/1721.1/101464 [accessed 2026-09-25]
55. Singh A, Brooks C, Wang X, Li W, Kim J, Wilson D. Bridging learnersourcing and AI: exploring the dynamics of student-AI collaborative feedback generation. Presented at: LAK '24: Proceedings of the 14th Learning Analytics and Knowledge Conference; Mar 18-22, 2024; Kyoto Japan p. 742-748. [doi: 10.1145/3636555.3636853]
56. Teng M, Singla R, Yau O, et al. Health care students' perspectives on artificial intelligence: countrywide survey in Canada. JMIR Med Educ 2022 Jan 31;8(1):e33390. [doi: 10.2196/33390] [Medline: 35099397]

## Abbreviations

**APA:** American Psychological Association
**AWS:** Amazon Web Services



XSL•FO
RenderX

**DP:** deliberate practice
**FAQ:** frequently asked question
**GPU:** graphics processing unit
**IRB:** institutional review board
**LLM:** large language model
**RQ:** research question
**TA:** teaching assistant

*Edited by M Balcarras; submitted 09.Feb.2026; peer-reviewed by A Hudon, T Yusuff; revised version received 25.Sep.2026; accepted 13.Oct.2026; published 11.Nov.2026.*

*Please cite as:*
*Converse E, Louie R, Safer D, Pacheco JP, Fang W, Kent J, Arnow B, Yang D*
*Classroom Integration of AI-Simulated Patients and Automated Feedback in Doctoral Psychotherapy Training: Longitudinal Mixed Methods Study*
*Interact J Med Res 2026;15:e93199*
*URL: https://www.i-jmr.org/2026/1/e93199*
*doi:10.2196/93199*

© Ellen Converse, Ryan Louie, Debra Safer, Juan Pablo Pacheco, William Fang, Jamie Kent, Bruce Arnow, Diyi Yang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 11.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Repeat At-Home Saliva Collection to Measure Cortisol Awakening Response (CAR): Protocol Feasibility Study Among Black Women With Histories of Abuse

Kiyomi Tsuyuki[1], MPH, PhD; Marguerite B Lucea[2], RN, MPH, PhD; Andrea N Cimino[3], MSW, PhD; Michael Killian[4], MSW, PhD; Christina J Catabay[1], MPH; Engle Abrams[3]; Emmanuel Baffour-Siaw[3]; Jacquelyn C Campbell[3], RN, BSN, MSN, PhD; Douglas A Granger[5,6], PhD; Jamila K Stockman[1], MPH, PhD

[1]Division of Infectious Diseases and Global Public Health, Department of Medicine, University of California, San Diego, 9500 Gilman Drive, Mail Code 0507, La Jolla, CA, United States

[2]Department of Nursing, College of Health Professions, Towson University, Towson, MD, United States

[3]Johns Hopkins University School of Nursing, Baltimore, MD, United States

[4]College of Social Work, Florida State University, Tallahassee, FL, United States

[5]Department of Psychology, University of California at Irvine, Irvine, CA, United States

[6]Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD, United States

**Corresponding Author:**
Kiyomi Tsuyuki, MPH, PhD
Division of Infectious Diseases and Global Public Health, Department of Medicine, University of California, San Diego, 9500 Gilman Drive, Mail Code 0507, La Jolla, CA, United States

## Abstract

**Background:** Emerging evidence supports that women with histories of abuse have heightened stress and immune dysregulation. Few studies have examined the biological plausibility of this association in US Black women—a population disproportionately affected by gender-based violence (GBV), health disparities, and generally underrepresented in research. Biomarkers of stress and immune health remain challenging to study due to barriers in recruitment, retention, and protocol adherence.

**Objective:** This study developed and examined the feasibility of an at-home, self-collected, and minimally invasive salivary cortisol awakening response (CAR) collection protocol among Black women with histories of abuse in Baltimore City, Maryland.

**Methods:** Black women were recruited from November 2015 to May 2018 from Baltimore City sexually transmitted disease clinics. Participants received in-person instruction and demonstration on saliva self-collection using the passive drool method, and were provided study-issued cell phones for reminders and adherence tracking. Participants self-collected saliva samples upon waking and 30 minutes post waking on 3 consecutive days to assess CAR. Outcomes included protocol completion, self-reported adherence, and experiences (via saliva collection logs), a weighted protocol adherence score, and cortisol measures (waking, post waking, and CAR). Differences in sociodemographic characteristics and outcome measures were assessed by adult forced sex exposure status using chi-square tests and ANOVA. The Pearson correlation coefficient assessed the day-to-day reproducibility of cortisol measures among women with 2 full days of valid samples.

**Results:** Of the 305 women completing the study survey, 228 completed the saliva specimen collection protocol, with no significant differences in completion between women with and without forced sex histories. Protocol feasibility was high, with 84% (191/228) completing at least 1 full day of adherent saliva collection and more than half (132/228) completing 2 full days. While feasibility did not differ by exposure status, women with adult forced sex exposure demonstrated lower protocol adherence scores compared to their unexposed counterparts. Between-day cortisol measures demonstrated variable reproducibility across waking, post waking, and CAR measures.

**Conclusions:** This at-home salivary cortisol collection protocol was feasible in our robust sample of Black women with histories of abuse. The process of at-home collection of salivary biomarkers, including cortisol measures, was well-documented, and participants were able to adhere to it. The protocol yielded usable data that will facilitate the examination of the physiological and health repercussions of GBV.

***(Interact J Med Res 2026;15:e64150)*** doi:10.2196/64150

**KEYWORDS**

saliva collection; black women; histories of abuse; gender-based violence; cortisol; stress biomarkers; immune dysregulation; protocol adherence



XSL•FO
RenderX

## Introduction

### Background

Chronic social stress is increasingly recognized as a key contributor to health disparities [1]. The minority stress model posits that individuals exposed to intersecting forms of social disadvantage experience cumulative stress that can be biologically embedded over time and adversely affect health [2]. Black women may face particularly high levels of chronic stress due to the intersection of racism, sexism, and a disproportionate burden of gender-based violence (GBV) [3,4]. Despite these intersecting stressors, few studies have incorporated biological measures of stress in research with women exposed to GBV. Existing studies often rely on small sample sizes, lack racial and ethnic diversity, collect clinic-based biomarkers, limiting participation and generalizability to the social context, and lack measures of collection protocol adherence [5]. Developing and testing a feasible and acceptable trauma-informed protocol for at-home, self-collection of stress biomarkers is a critical step toward advancing health equity research on the physiological impacts of violence. In this paper, we establish a protocol of repeat salivary cortisol self-collection among a robust sample of Black women with experiences of violence and examine the feasibility and usability of the samples.

GBV is a pervasive public health problem. Globally, an estimated 30% - 60% of women experience physical and/or sexual violence during their lifetime [6]. In the United States, 42% of women experience physical intimate partner violence (IPV) and 20% experience sexual violence in their lifetime [7]. Black women experience disproportionately high rates of IPV, including physical, sexual, and stalking IPV, compared with their White and Latina counterparts [7]. The health consequences of GBV are substantial. Survivors of IPV are more likely to experience mental health conditions (eg, depression, anxiety, posttraumatic stress disorder, and suicidality) [7,8] and chronic physical conditions, including cardiovascular disease, chronic pain, gastrointestinal disorders, sexual and reproductive health disorders, and increased susceptibility to infectious diseases, than women without IPV experiences [9]. Population-based research further suggests that Black women survivors of severe IPV demonstrate higher rates of mental health disorders and report poorer overall health than their White counterparts [10].

Cortisol is a widely used biomarker of physiological stress. Researchers have measured cortisol levels among women with histories of abuse using plasma [11,12], hair [13], and saliva [11,14,15]. However, this research has yielded inconsistent findings associated with cortisol levels in the context of IPV, likely due to variation in measurement approaches and cortisol analysis [13]. Salivary cortisol has emerged as a valuable tool because it is minimally invasive, convenient, and well-suited for repeat sampling. Saliva collection enables participants to self-collect samples in naturalistic settings (eg, at home), facilitating the study of hypothalamic-pituitary-adrenal (HPA) axis activity within the context of their daily life rather than in clinical settings. This approach further supports the inclusion of larger and more diverse samples, including populations that may face barriers to participation in clinic-based research. However, salivary cortisol research among women with abuse histories has been limited by several methodological challenges, including insufficient attention to collection adherence measures, lack of racially and ethnically diverse samples, reliance on clinic-based collection, and inadequate sampling frequency within or across days, which is critical to accurately quantify cortisol levels and patterns [16].

Cortisol follows a diurnal rhythm, including a rapid increase within 30 - 45 minutes after waking, known as the cortisol awakening response (CAR) [17]. Alterations in the CAR have been linked to chronic stress exposure and a range of adverse health outcomes disparities [18,19], making it a critical index in stress research. Accurate assessment of CAR depends on precise timing of saliva collection immediately upon waking and at specific postwaking intervals [20,21], introducing challenges for at-home saliva self-collection protocol adherence [22]. Although previous work has attempted to standardize CAR measurement, issues of protocol adherence remain a significant concern, particularly in unsupervised settings. These challenges may be amplified in research among women with histories of abuse, for whom ethical, safety, and trauma-informed considerations are critical. No standardized protocol exists for at-home salivary CAR self-collection for women with histories of abuse.

Medical mistrust is a critical barrier to participation in research for women with abuse histories and among Black women, and must be addressed with trauma-informed practices. Women with abuse histories often distrust medical institutions due to previous negative experiences, fear of judgment, and concerns about confidentiality [23,24]. Gender and abuse biases within medical settings can lead to the dismissal of women's health concerns, while systemic failures in addressing abuse further exacerbate mistrust and discourage care-seeking. Among Black women, medical mistrust is rooted in both historical and ongoing discrimination within the health care system [25,26]. Black women have been disproportionately subjected to unethical medical practices, such as forced sterilization and nonconsensual experimentation, and report higher levels of perceived discrimination in health care settings [25]. Medical mistrust is linked to poorer health outcomes, contributes to health disparities, and likely reduces willingness to participate in research and specimen collection [27-30]. Addressing these barriers requires the implementation of trauma-informed care approaches. Developing an at-home, repeated measures saliva collection protocol requires attending to these barriers to ensure inclusive participation.

### Objectives

The objectives of this study are 2-fold. First, we describe the at-home, salivary CAR self-collection protocol designed for Black women with histories of abuse. Second, we assess the feasibility of this protocol among Black women with histories of abuse.



XSL•FO
RenderX

## Methods

### Study Design

Data were drawn from the ESSENCE (Examining Stress, Sexual Experiences, and Neighborhood Correlates of HIV Risk among Black Women) project, a retrospective cohort study to examine how experiences of forced sex and physiological stress response contribute to HIV risk among Black women. This study focuses on the development and implementation of an at-home saliva CAR self-collection protocol embedded within the broader study.

### Study Setting

The study was conducted in 2 publicly funded sexually transmitted disease (STD) clinics in Baltimore, Maryland, United States, from November 2015 to May 2018. Participants were recruited from neighborhoods that have historically experienced structural inequities, such as racism, residential segregation, and community disinvestment [31], as well as unethical research practices surrounding Henrietta Lacks that spurred medical mistrust in the communities [32]. Data collection occurred during a period of heightened community tension following the death of Freddie Gray in police custody in April 2015 and the subsequent criminal proceedings that ultimately exonerated the 6 officers who were involved in July 2016, and occurred in neighborhoods where the protests were most intense [33]. These contextual factors informed the study's trauma-informed and community-engaged approach to participant recruitment and saliva collection protocol development.

### Recruitment and Eligibility

Black women seeking health services were recruited in the clinic waiting rooms by trained female research staff. To prioritize safety and confidentiality, staff discreetly approached women and invited them to participate in an anonymous study about "women's health and safety." Interested women were led to a private examination room where staff read the informed consent document.

After providing informed consent, participants responded to a short eligibility screening survey. Eligibility criteria included (1) biologically female, (2) between ages 18 - 44 years, (3) self-identified as African American or Black, (4) reported having sex with a man in the past 6 months, and (5) reported having either 2 or more sexual partners in the past year or having a sexual partner at high risk for HIV (ie, used injection or noninjection drugs, had sex with men, been to prison, had a concurrent sex partner, had an STD, was HIV-positive, or did not know if their sexual partner had any of these characteristics).

### Ethical Considerations

The study protocol was approved by the institutional review boards of Johns Hopkins University and the University of California, San Diego. Participants provided written informed consent separately for the eligibility screening and study participation, including the survey and saliva collection components. Although the study collected geographic data and contact information for follow-up, confidentiality was maintained by storing consent forms and identifiable information separately from the survey responses and biomarker data. All study data were deidentified and linked to a participant ID. Participants received financial compensation totaling US $70 (US $10 for screening, US $25 for survey completion, and US $35 for returning saliva samples), a daily bus pass, and a list of community resources related to violence and health care service referrals.

### Survey Data Collection

Participants completed a 60 - to 90-minute survey administered using audio computer-assisted self-interview software. The survey assessed sociodemographic characteristics, sexual health and behavior, violence and trauma history, mental health, substance use, and neighborhood and social environmental factors. A female research staff member was made available to the participants in case of any questions, concerns, or technical difficulties. Following survey completion, participants were trained on saliva collection procedures [34].

### Salivary Cortisol Collection Protocol

#### *In-Clinic Training and Demonstration*

##### Overview

After completing the survey, participants received standardized training on saliva collection procedures from research staff using a scripted protocol. Training covered the passive drool technique, saliva collection timing to measure the CAR, adherence documentation procedures, sample storage, and sample retrieval process. Participants then observed a demonstration of the passive drool technique and were given the opportunity to practice. Participants then received a collection kit containing 6 straws, 6 barcoded cryovials (with color-coded caps to visually distinguish waking [green] and 30-minute postwaking [red] samples), a saliva collection log (with adherence instructions on the back), a pen, and a study-issued mobile phone with charger. Study staff used the mobile phones to maintain contact with participants during the collection period to provide reminders, troubleshoot issues, and coordinate sample retrieval.

##### Passive Drool Technique

Saliva samples were collected using the passive drool method developed by Salimetrics [34]. This method collects whole saliva rather than gland-specific saliva obtained from oral swabs. Participants generated saliva by thinking about a favorite food and generating saliva by gently moving their jaws as if chewing and allowing saliva to pool in the mouth. Using a straw attached to a cryovial, participants then gently allowed the saliva to flow into the vial until the 1 mL fill-line was reached.

##### CAR Collection

Participants were instructed to collect 2 saliva samples per day for 3 consecutive days (6 samples in total)—one sample immediately upon waking and a second sample 30 minutes post waking [35]. This schedule allowed the measurement of the CAR. Participants were instructed to avoid brushing their teeth, smoking, or eating before completing each morning's saliva collection.



XSL•FO
RenderX

### Adherence Monitoring

The saliva collection protocol used multiple adherence monitoring strategies to measure the timing of saliva collection, as the accurate measurement of CAR requires precise timing of saliva collection relative to waking time. Participants were asked to record collection times using (1) a saliva collection log where wake times and saliva collection times were written in the log (Multimedia Appendix 1), (2) time-stamped photos of each saliva collection cryovial using the study-issued phone, and (3) text message responses to automated SMS text message prompts confirming sample collection times.

The saliva collection log included saliva protocol instructions, vial ID numbers, and space to document any stressors or other issues that may have impacted saliva collection. Study staff recorded participants' anticipated wake times and expected collection times for each study day on the saliva collection log. In the saliva collection log, participants were further instructed to report any mistakes or protocol deviations during saliva collection, as well as stressors experienced the previous night.

SMS text message reminders were delivered via an automated SMS digital platform and were scheduled based on participants' anticipated wake times. Messages prompted participants to confirm when they woke up and when they began collecting both the waking sample and the 30 minutes postwaking sample. These procedures were repeated across all 3 collection days.

### Sample Storage

After each sample was collected, participants were instructed to place the vial in their home freezer until study staff retrieved the samples.

## Sample Retrieval and Field Procedures

After completing the 3-day saliva collection period, study staff retrieved the frozen samples, saliva collection logs, and study phones from the participants' homes. Study staff also administered a brief survey assessing participants' understanding and adherence to the saliva collection protocol and reporting the study-issued phone condition (Multimedia Appendix 2). Participants received compensation (US $35) upon sample return.

Before the saliva retrieval home visit, study staff confirmed appointment times and assessed potential safety considerations. When safety concerns were present, alternative sample retrieval locations were arranged. Study staff documented any safety concerns in participant files, including a codeword that the participants could use to alert staff that they were unsafe and should call 911. Moreover, 2 study staff were present to conduct the retrieval visits, typically conducted on weekdays during daytime hours. Exceptions were made on a case-by-case basis for saliva sample retrieval visits during the weekend or at a location other than the participant's home (ie, their workplace or a friend's home).

At retrieval, study staff visually inspected each vial to verify the presence of at least 1 mL frozen saliva and recorded vial ID numbers, saliva collection information, and retrieval visit details. Study staff also administered a brief survey assessing participants' understanding and adherence to the saliva collection protocol. Samples were then transported in a small cooler with ice packs to a laboratory deep freezer at Johns Hopkins University, where they were stored at −80 °C until shipped out for assay.

## Trauma-Informed Protocol

The saliva self-collection protocol was developed using trauma-informed principles to enhance feasibility, acceptability, and adherence among women with histories of abuse. Standardized procedures emphasized safety, autonomy, and privacy by allowing self-collection at home. Study materials were scripted with nonstigmatizing, culturally sensitive language and included explicit statements of voluntary participation, permitting participants to pause or discontinue at any time. Although collection schedules were standardized, the protocol allowed for flexible collection schedules, and reminders were delivered using preapproved methods. Participants were also provided with community and psychosocial support resources. Collectively, these procedures were designed to minimize retraumatization while supporting protocol compliance.

## Field Coordination

A full-time field coordinator oversaw saliva collection procedures, maintained participant contact, and coordinated sample retrieval. The field coordinator was a Black woman who was a member of the local community, which facilitated trust and communication with participants and allowed her to understand the participants' lived experiences. Participants responded positively to the field coordinator's one-on-one supportive phone calls. This role was critical to providing support for adherence to the saliva collection protocol.

The field coordinator provided great insight into the facilitators and barriers to study participation and to saliva collection, helping to refine the saliva collection protocol in several instances. First, the protocol originally specified a collection schedule based on recruitment date (ie, those recruited into the study on a Monday would collect saliva on Tuesday through Thursday). The field coordinator reported that a predetermined saliva collection start schedule did not fit many participants' life demands. The protocol was adjusted to allow for a flexible collection start date. Second, the field coordinator was flexible to changes in call or text times and saliva retrieval schedules, often working outside of typical business hours to meet participants' unique needs and facilitate their study participation. This support and much-needed flexibility allowed participants to better adhere to the saliva collection protocol. Third, the field coordinator conducted in-home saliva collection demonstrations for participants who needed the extra support, adding further support for protocol adherence.

## Sample Selection for Analysis

After saliva samples were retrieved, the 2 collection days most adherent to the protocol were chosen for shipping, assaying, and analysis. We developed a decision tree algorithm (Multimedia Appendix 3) to identify the 2 most protocol-adherent collection day samples to be assayed. Protocol adherence was evaluated using time data from saliva logs, time-stamped photographs, and SMS text messages. The algorithm prioritized samples collected closest to the target



XSL•FO
RenderX

timing (eg, immediately upon waking and 30-min post waking) and gave greater weights to samples in which collection time was only manually noted on the saliva log, and for every 5-minute delay in saliva collection. The 2 most adherent collection days (with the lowest total weight) were selected for assay.

### Sample Storage and Shipping

Saliva samples were frozen immediately after collection and kept frozen for analysis. Samples retrieved from participants in Baltimore, Maryland, were stored at –80 °C at Johns Hopkins University until shipment and then batch shipped overnight on dry ice in accordance with federal regulations for biological materials to the Institute of Interdisciplinary Salivary Bioscience Research at the University of California, Irvine.

### Cortisol Assay

At the Institute of Interdisciplinary Salivary Bioscience Research, on the day of assay, samples were thawed and centrifuged to remove mucins and other particulates. Cortisol concentrations were measured in duplicate using a commercially available enzyme immunoassay (Salimetrics). The assay had a lower detection limit of 0.007 µg/dL, a range up to 3.0 µg/dL, and average inter- and intra-assay coefficients of variation below 10% and 5%, respectively [36]. The average of duplicate assay results for each sample was used in the statistical analyses.

### Measures

*Outcome variables* included (1) completing the saliva collection protocol, (2) saliva collection log, (3) saliva collection protocol adherence weight, and (4) cortisol levels. *Completing the saliva collection protocol* included women who provided at least 1 saliva sample that was sent off for assay. *Saliva collection log* survey measured participants' adherence, understanding, and experiences with the at-home saliva collection protocol. Items captured compliance with instructions (eg, reviewing logs and following precollection restrictions), the usefulness of study reminders (eg, SMS text messages), assessed challenges (eg, remembering collection times and storing samples), as well as documented field notes. *Saliva collection protocol adherence weight* was a weighted score derived from the decision tree algorithm. For each collection day, weights (range 6 - 14 points) were assigned to each of the 2 samples collected (ie, waking sample and 30-min postwaking sample) based on the method of time verification (eg, time-stamped photo or text vs manual log) and adherence to the protocol time sampling windows, with penalties applied in 5-minute increments for delays. The 2 weights were summed for a total daily sample adherence weight (range 12 - 28 points). Lower scores indicated greater adherence to the protocol. *Cortisol levels* were reported as waking cortisol levels (µg/dL), 30-minute postwaking cortisol levels (µg/dL), and CAR. CAR was calculated by the area under the curve between cortisol measures upon waking and 30 minutes post waking using the following equation: CAR=((Cortisol 30-min post-waking+cortisol waking)/2)×T1)–(cortisol waking×T1), where T1 represents time elapsed between sample collections.

*Other variables* included (1) adult forced sex exposure and (2) sociodemographic variables. *Adult forced sex exposure* status was measured as a positive response to one of the following questions: (1) "Since you turned 18, has a male sex partner: used threats to make you have sex when you did not want to or used force (like hitting, holding down, or using a weapon) to make you have sex?" and (2) "Since you turned 18, has any other male done any of the following: used threats to make you have sex when you did not want to or used force (like hitting, holding down, or using a weapon) to make you have sex?" *Socio-demographic variables* included age, education, employment status, housing instability, annual individual income, relationship status, number of children, and sexual orientation.

### Statistical Analysis

All statistical analyses were conducted using Stata (version 15; StataCorp) [37]. Of the 305 women who participated in the study with complete survey data, 228 women completed saliva collection, with 191 (84%) women completing at least 1 full day of saliva collection that fell within acceptable range of time collection for CAR measurement (eg, waking sample collected within 15 min of wake time, 30-min postwaking sample collected between 30 - 45 min postwake time, and ≥20 min time difference between collection of waking and postwaking samples; data not shown).

We began our analysis with descriptive statistics of the sociodemographic characteristics of women in the overall sample who completed the saliva collection protocol. Chi-square statistic was used for categorical variables, and the *F* statistic from ANOVA was used for continuous variables to detect statistically significant differences by adult sexual violence exposure, using $P<.05$ for statistical significance cutoff. We then described the frequency distribution of responses to the saliva collection log, which gave insight into the feasibility and use of supports for adherence to the saliva collection protocol, and compared differences by adult sexual violence exposure. We then described the protocol adherence feasibility and reported the average sample adherence weights for the total sample and by adult sexual violence exposure. Finally, we reported the salivary cortisol levels for the total sample and by adult sexual violence exposure for the waking and 30-minute postwaking samples, and the CAR for the subsample of women who completed at least 2 full days of saliva collection within acceptable time ranges. We used the Pearson correlation coefficient to assess the significant correlation of cortisol levels between collection days.

## Results

### Overview

Overall, our results report on our saliva collection protocol results, and demonstrate the feasibility of saliva CAR self-collection among Black women with histories of abuse. Of the 305 women who were eligible and consented to the survey and saliva collection, 228 (75%) completed the saliva specimen collection protocol, 33 (11%) were lost at the survey, 23 (8%) refused the saliva collection, and 21 (7%) did not complete the saliva collection (refused mid-completion; data not shown). Of the women who completed the saliva specimen collection protocol, 93 (41%) had a history of forced sex since age 18



XSL•FO
RenderX

years (exposed women) and 135 (59%) had no history of forced sex since age 18 years (unexposed women).

Table 1 describes the demographic characteristics of the total study sample that completed the saliva collection protocol and compares differences by adult forced sex exposure. Among the total sample of Black women who completed the saliva collection protocol (n=228), participants were on average 26.6 (SD 6.45) years old, most had a high-school education or more (190/228, 83%), and over half were currently employed (133/228, 58%). Socioeconomic vulnerability was evident, with 61% (139/228) reporting annual incomes below US $10,000 and 7% (16/228) reporting housing instability. Most participants were single (160/228, 70%), had at least 1 child (118/228, 52%), and identified as heterosexual (192/228, 84%). Women exposed to adult sexual violence who completed the saliva collection were, on average, significantly older, were less likely to have a high-school education, were less likely to be currently employed, and were more likely to report housing instability than women unexposed to adult sexual violence who completed the saliva collection.

**Table .** Sociodemographic characteristics of Black women with violence experience completing repeated measures self-collection of saliva in Baltimore, Maryland (Source: The ESSENCE Project 2015-2018).

| Characteristic | Total saliva (n=228), n (%) | Adult forced sex exposure status | | |
|---|---|---|---|---|
| | | Exposed (n=93) | Unexposed (n=135) | *P* value[a] |
| Age[a] (y), mean (SD) | 26.61 (6.45) | 28.75 (0.73) | 25.13 (0.48) | <.001 |
| 18 - 24, n (%) | 96 (42) | 26 (28) | 70 (52) | |
| 25 - 34, n (%) | 102 (45) | 47 (51) | 55 (41) | |
| 35 - 44, n (%) | 30 (13) | 20 (22) | 10 (7) | |
| High-school education or more, n (%) | 190 (83) | 72 (77) | 118 (87) | .047 |
| Currently employed, n (%) | 133 (58) | 45 (48) | 88 (85) | .01 |
| Housing instability, n (%) | 16 (7) | 13 (14) | 3 (2) | .001 |
| Annual individual income (US $), n (%) | | | | .27 |
| <10,000 | 139 (61) | 62 (67) | 77 (57) | |
| 10,000 - 29,999 | 71 (31) | 26 (28) | 45 (33) | |
| ≥30,000 | 18 (8) | 5 (5) | 13 (10) | |
| Relationship status, n (%) | | | | .40 |
| Single | 160 (70) | 63 (68) | 97 (72) | |
| In relationship | 62 (27) | 26 (28) | 36 (27) | |
| Separated, divorced, widowed, or other | 6 (3) | 4 (4) | 2 (1) | |
| Number of children[a], mean (SD) | 1.07 (1.32) | 1.12 (0.14) | 1.04 (0.11) | .68 |
| 0, n (%) | 110 (48) | 44 (47) | 66 (49) | .91 |
| 1 to 2, n (%) | 84 (37) | 34 (37) | 50 (37) | —[b] |
| 3 or more, n (%) | 34 (15) | 15 (16) | 19 (14) | — |
| Sexual orientation, n (%) | | | | .007 |
| Heterosexual | 192 (84) | 71 (76) | 121 (90) | |
| Bisexual or gay | 36 (16) | 22 (24) | 14 (10) | |

[a]*P* value for statistical difference by adult forced sex exposure status using chi-square for comparisons among categorical variables and the *F* statistic for comparison of means among continuous variables.

[b]Not applicable.

## Saliva Collection Logistics

Analysis of the saliva collection log data (Table 2) demonstrates high levels of protocol feasibility, regardless of adult forced sex exposure. In terms of the saliva collection demo, most women reported receiving the saliva demo in the clinic 192/228, 84%), most reported that the saliva demo at the clinic was thorough enough to remember how to self-collect (190/228, 83%), and 99% (225/228) returned the saliva log, which was the standard protocol. In terms of protocol adherence supports, most women



XSL•FO
RenderX

reported referring to the instructions page provided to help remember how to collect the saliva sample (160/228, 70%). The most difficult part of the saliva collection reported was drooling (74/228, 33%) followed by taking pictures of the sample (46/228, 20%). Most women reported having a little or moderate difficulty refraining from drinking, eating, smoking, or brushing their teeth during the saliva collection (125/228, 55%). In terms of the study phones, most women reported receiving texts from the study phone (206/228, 90%), with most reporting that these texts were moderately useful (131/228, 58%) or very useful (56/228, 25%), and most using the study phone texts to wake up to collect their first sample (140/228, 61%). The majority of women reported taking pictures of the saliva vials (211/228, 93%). In terms of saliva storage, most women reported no problems storing the saliva samples in their freezer (217/228, 95%). There were no statistically significant differences observed between women exposed and unexposed to adult forced sex across the saliva collection log measures, indicating that the protocol was feasible across exposure groups.



XSL•FO
RenderX

**Table .** Saliva self-collection logistics among Black women with violence experience, by adult forced sex exposure status (n=228), Baltimore, Maryland (Source: The ESSENCE Project 2015-2018).

| Characteristic | Saliva total (n=228), n (%) | Adult forced sex exposure status | | |
|---|---|---|---|---|
| | | Exposed (n=93), n (%) | Unexposed (n=135), n (%) | *P* value[a] |
| Saliva demo | | | | |
| Saliva demo location | | | | —[b] |
| In clinic | 192 (84) | 76 (82) | 116 (86) | |
| In home | 1 (0) | 1 (1) | 0 (0) | |
| In office | 30 (13) | 13 (14) | 17 (13) | |
| Missing | 5 (2) | 3 (3) | 2 (1) | |
| Saliva demo at clinic was thorough enough to remember how to collect | | | | .09 |
| Yes | 190 (83) | 72 (77) | 118 (87) | |
| No | 26 (11) | 13 (14) | 13 (10) | |
| Missing | 12 (5) | 8 (9) | 4 (3) | |
| Saliva log | | | | |
| Returned the saliva log | | | | .36 |
| Yes | 225 (99) | 91 (98) | 134 (99) | |
| No | 3 (1) | 2 (2) | 1 (1) | |
| Protocol adherence supports | | | | |
| Referred to saliva instructions page to help remember how to collect | | | | .22 |
| Yes | 160 (70) | 71 (76) | 89 (66) | |
| No | 57 (25) | 19 (20) | 38 (28) | |
| Missing | 11 (5) | 3 (3) | 8 (6) | |
| Most difficult part of saliva collection | | | | .42 |
| Drooling | 74 (32) | 35 (38) | 39 (29) | |
| Taking pictures | 46 (20) | 19 (20) | 27 (20) | |
| Remember when to collect sample | 30 (13) | 12 (13) | 18 (13) | |
| Replying to texts | 20 (9) | 8 (9) | 12 (9) | |
| Logging info | 9 (4) | 1 (1) | 8 (6) | |
| Missing | 49 (21) | 18 (19) | 31 (23) | |
| Difficult to refrain from drinking, eating, smoking, or brushing teeth | | | | .20 |
| Not at all difficult | 54 (24) | 16 (17) | 38 (28) | |
| A little or moderately difficult | 125 (55) | 52 (56) | 73 (54) | |
| Difficult | 27 (12) | 15 (16) | 12 (9) | |
| Very difficult | 6 (3) | 2 (2) | 4 (3) | |
| Missing | 16 (7) | 8 (9) | 8 (6) | |
| Study phones | | | | |
| Received texts from study phone | | | | .61 |
| Yes | 206 (90) | 86 (92) | 120 (89) | |
| No | 17 (7) | 5 (5) | 12 (9) | |
| Missing | 5 (2) | 2 (2) | 3 (2) | |
| Study phone texts were very useful in helping me collect saliva | | | | .75 |
| Not at all useful | 38 (17) | 13 (14) | 25 (19) | |



XSL•FO
RenderX

| Characteristic | Saliva total (n=228), n (%) | Adult forced sex exposure status | | |
|---|---|---|---|---|
| | | Exposed (n=93), n (%) | Unexposed (n=135), n (%) | *P* value[a] |
| Moderately useful | 131 (57) | 57 (61) | 74 (55) | |
| Very useful | 56 (25) | 22 (24) | 34 (25) | |
| Missing | 3 (1) | 1 (1) | 2 (1) | |
| Used study phone texts to wake up to collect your sample | | | | .16 |
| Yes | 140 (61) | 64 (69) | 76 (56) | |
| No | 73 (32) | 24 (26) | 49 (36) | |
| Missing | 15 (7) | 5 (5) | 10 (7) | |
| Took pictures of the saliva vials | | | | .50 |
| Yes | 211 (93) | 84 (90) | 127 (94) | |
| No | 14 (6) | 7 (8) | 7 (5) | |
| Missing | 3 (1) | 2 (2) | 1 (1) | |
| Saliva storage | | | | |
| Any problems storing the samples in the freezer | | | | .24 |
| Yes | 8 (4) | 1 (1) | 7 (5) | |
| No | 217 (95) | 91 (98) | 126 (93) | |
| Missing | 3 (1) | 1 (1) | 2 (1) | |

[a]*P* value for statistical difference between adult forced sex exposure status using $\chi^2$ for comparisons.

[b]Not applicable.

Table 3 describes the feasibility of protocol completion and assesses protocol adherence using the protocol adherence weights generated by the decision tree algorithm. Protocol completion was highly feasible, with the majority (191/228, 84%) of the total sample successfully completing at least 1 full day of saliva collection within the specific timing necessary to measure CAR. Of the total sample, 26% (59/191) achieved 1 full day of adherence to protocol collection, and 58% (132/191) achieved 2 full days of adherence. This was consistent across adult forced sex exposure categories, with 83% (77/93) of exposed women and 84% (114/135) of unexposed women completing at least 1 full day of adherent saliva collection, with no significant differences. In terms of protocol adherence weights, the total sample maintained an excellent mean adherence weight of 16.25 (SD 6.303). There were statistically significant differences in the total protocol adherence weights by adult forced sex exposure status, with exposed women demonstrating significantly less adherence to the protocol (mean 17.85, SD 5.217) than unexposed women (mean 15.16, SD 6.750; *P*=.002). This discrepancy was most pronounced among women who completed the full 2-day protocol, where exposed women remained significantly less adherent to the protocol than their unexposed counterparts (mean 17.36, SD 4.986 vs mean 14.48, SD 6.126, respectively; *P*=.004). This finding suggests that, although feasibility was high across both groups, trauma history may influence women's ability to adhere to the protocol.



XSL•FO
RenderX

**Table .** Compliance to saliva self-collection protocol using adherence weights for Black women with violence experience completing repeated measures, by adult forced sex exposure status, Baltimore, Maryland (Source: The ESSENCE Project 2015-2018).

| Protocol feasibility | Saliva total | | Adult forced sex exposure status | | | | |
|---|---|---|---|---|---|---|---|
| | n (%) | Mean (SD) | Exposed | | Unexposed | | *P* value[a] |
| | | | n (%) | Mean (SD) | n (%) | Mean (SD) | |
| Total samples collected within time range | 228 (100) | —[b] | 93 (100) | — | 135 (100) | — | .88 |
| 0 full day | 37 (16) | — | 16 (17) | — | 21 (16) | — | |
| 1 full day | 59 (26) | — | 25 (27) | — | 34 (25) | — | |
| 2 full days | 132 (58) | — | 52 (56) | — | 80 (59) | — | |
| Sample adherence weight | | | | | | | |
| Total sample | 191 (100) | 16.25 (6.303) | 77 (100) | 17.851 (5.217) | 114 (100) | 15.164 (6.750) | .002 |
| 1 full day | 59 (31) | 17.669 (7.048) | 25 (32) | 18.880 (5.632) | 34 (30) | 16.779 (7.894) | .24 |
| 2 full days | 132 (69) | 15.612 (5.857) | 52 (68) | 17.356 (4.986) | 80 (70) | 14.478 (6.126) | .004 |

[a]Of samples with adherence weights; *P* value for statistical difference between Exposed/Unexposed using Chi-square for comparisons among categorical variables & F-statistic for comparison of means among continuous variables.

[b]Not applicable.

Table 4 reports the cortisol levels among women who completed 2 full days of saliva collection within the required time limits for CAR, comparing the Pearson correlation coefficient of between-day cortisol levels (waking, post waking, and CAR) for each of the 2 days. Among the total sample, day 1 and day 2 cortisol waking samples ($r$=0.276; $P$=.001) and the CAR levels ($r$=0.178; $P$=.04) were weakly but statistically significantly correlated, whereas the postwaking samples ($r$=0.411, $P$<.001) were moderately and statistically significantly correlated. When stratifying by adult forced sex exposure, waking cortisol samples for day 1 and day 2 were weakly but statistically significantly correlated among exposed women ($r$=0.316, $P$=.02) and unexposed women ($r$=0.258, $P$=.02). Postwaking cortisol samples for day 1 and day 2 were moderately and significantly correlated among exposed women ($r$=0.505, $P$<.001) and unexposed women ($r$=0.380, $P$=.001). CAR levels for day 1 and day 2 were not statistically correlated among exposed women ($r$=0.027, $P$=.85); however, CAR levels were weakly but statistically significantly correlated among unexposed women ($r$=0.273, $P$=.01). Overall, postwaking cortisol demonstrated the greatest reproducibility across the 2 days, while CAR demonstrated the least stability, particularly among women with adult forced sex exposure.



XSL•FO
RenderX

**Table .** Salivary cortisol levels for Black women with violence experience completing 2 days of adherent saliva self-collection, by adult forced sex exposure status, Baltimore, Maryland (Source: The ESSENCE Project 2015-2018).

| | Total | | Adult forced sex exposure status | | | | *P* value[a] | |
|---|---|---|---|---|---|---|---|---|
| | | | Exposed | | Unexposed | | Sample 1 | Sample 2 |
| | Sample 1 | Sample 2 | Sample 1 | Sample 2 | Sample 1 | Sample 2 | | |
| Waking[b] | | | | | | | .75 | .44 |
| Women, n | 132 | 132 | 52 | 52 | 80 | 80 | | |
| Mean (SD) | 0.29 (0.26) | 0.24 (0.17) | 0.30 (0.25) | 0.25 (0.23) | 0.29 (0.26) | 0.23 (0.13) | | |
| Median (IQR) | 0.21 (0.131 to 0.366) | 0.21 (0.125 to 0.289) | 0.21 (0.126 to 0.415) | 0.21 (0.131 to 0.280) | 0.21 (0.134 to 0.362) | 0.22 (0.124 to 0.308) | | |
| Post waking[c] | | | | | | | .82 | .55 |
| Women, n | 132 | —[d] | 52 | — | 80 | — | | |
| Mean (SD) | 0.37 (0.30) | 0.34 (0.21) | 0.36 (0.25) | 0.35 (0.17) | 0.37 (0.33) | 0.33 (0.23) | | |
| Median (IQR) | 0.30 (0.175 to 0.296) | 0.32 (0.168 to 0.457) | 0.29 (0.199 to 0.517) | 0.35 (0.218 to 0.463) | 0.31 (0.173 to 0.514) | 0.28 (0.165 to 0.433) | | |
| CAR[e,f] | | | | | | | .42 | .68 |
| Women, n | 132 | — | 52 | — | 80 | — | | |
| Mean (SD) | 1.287 (3.20) | 1.637 (2.95) | 1.007 (3.15) | 1.504 (3.01) | 1.469 (3.24) | 1.723 (2.93) | | |
| Median (IQR) | 0.822 (−0.405 to 2.738) | 1.137 (0.0675 to 2.948) | 0.750 (−0.473 to 2.675) | 1.419 (0.030 to 3.109) | 0.867 (−0.402 to 2.813) | 0.780 (0.122 to 2.666) | | |

[a]*P* value for statistical difference between exposed or unexposed using *F* statistic for comparison of means among continuous variables.

[b]Total: $r$=0.276, $P$=.001; exposed: $r$=0.316, $P$=.02; unexposed: $r$=0.258, $P$=.02.

[c]Total: $r$=0.411, $P$<.001; exposed: $r$=0.505, $P$<.001; unexposed: $r$=0.380, $P$=.001.

[d]Not applicable.

[e]CAR: cortisol awakening response.

[f]Total: $r$=0.178, $P$=.04; exposed: $r$=0.027, $P$=.85; unexposed: $r$=0.273, $P$=.01.

# Discussion

## Principal Findings

The study developed a standardized, trauma-informed protocol for multiple day at-home salivary CAR self-collection for women with histories of abuse and provides evidence supporting the overall feasibility of implementing the protocol among a hard-to-reach sample of US Black women with experiences of trauma and abuse.

Overall participants demonstrated high engagement with study procedures, with the majority returning saliva logs, using adherence supports (ie, SMS text message reminders and referring to instructional materials), and adhering to key protocol components (ie, documenting sample collection times and freezing samples). However, findings highlight critical differences in protocol feasibility and adherence among women who completed the full 2-day collection by adult sexual violence exposure status, with exposed women demonstrating lower protocol adherence scores than unexposed women. Future protocols may benefit from enhanced trauma-informed supports, such as more frequent reminders, simplified procedures, or real-time monitoring, to mitigate adherence challenges among this population.

The cortisol analyses provide additional insight into the reliability of the at-home self-collected salivary biomarkers. Postwaking cortisol levels demonstrated the strongest reproducibility across days, while CAR demonstrated weaker and less consistent correlations, especially among women exposed to adult forced sex. The lack of significant day-to-day correlation in CAR among exposed women may reflect both biological variability with trauma-related HPA axis dysregulation and the observed challenges with protocol adherence. These findings suggest that while home-based saliva self-collection for CAR measurement is a viable and inclusive approach for studying stress physiology in this population, careful attention to protocol adherence and collection day flexibility is warranted.

This study is innovative in developing a saliva collection protocol tailored for this underrepresented population. Best practices include the following: (1) the use of text reminders via cell phones to enhance participants' adherence to the saliva collection protocol; (2) the use of study staff who were representative of the community to retrieve saliva samples; and (3) the creation of a saliva sample decision tree, which automated the decision of which 2 of 3 saliva samples were most adherent in such a large sample. Research implications include the facilitation of the collection of salivary biomarkers (including stress and inflammation measures) to promote



XSL•FO
RenderX

research that examines the physiological and health repercussions of GBV within the context of women's everyday lives.

### Strengths and Limitations

The protocol had several strengths. First, we established an at-home saliva self-collection (and storage) protocol with accommodations for women currently experiencing violence. Second, we had a large sample size of Black women with violence experience to demonstrate feasibility and acceptability in a diverse sample with safety considerations. Our protocol, importantly, demonstrated feasibility and acceptability in 2 populations (Black women and Black women with abuse histories) that have classically been mistrustful of the health care system and health care research. Third, we explored the use of study-issued cell phones to improve protocol adherence in terms of timing of saliva collection (as CAR is time-sensitive) and feasibility of sending a time-stamped picture documenting the time of the collected sample. Finally, collecting saliva at-home captures the interaction between biological and behavioral processes in the context of participants' everyday life, potentially enhancing the information collected compared with collecting biological markers in clinical settings.

Our protocol had several limitations that researchers implementing similar studies may wish to address. First, documenting the timing of saliva collection using the standard written log may have introduced self-report bias. To mitigate this, participants were instructed to send a time-stamped photo immediately after collecting each saliva sample. However, adherence to this protocol procedure was inconsistent due to logistical barriers (eg, uncharged or inaccessible study phone). Future studies could improve data accuracy by implementing automated time-stamping technologies (eg, mobile apps and electronic monitoring devices). Second, establishing regular wake times for participants for 3 consecutive days was challenging, particularly for participants who worked nights or with irregular sleep schedules. Although participants were provided with standardized instructions to plan sample collection based on their primary wake time, wake time ambiguity persisted among participants with fragmented sleep patterns. This variability may have introduced "noise" in terms of timing of sample collection relative to the cortisol diurnal pattern. Future research should consider more flexible, individualized collection protocols for women with abuse histories. Third, study-issued phones were intended to reduce the associated cost and participant burden of using their personal phone. However, some participants did not consistently use the study-issued phones, misused them for personal purposes, or reported loss or damage. Most participants had their own cell phone, and it was difficult for them to carry and use 2 cell phones. Additionally, for participants with histories of abuse, carrying a secondary phone may have posed safety concerns (ie, raising the suspicions of an abuser), which potentially affected protocol adherence. Future studies should carefully weigh the risks and benefits of providing study-specific devices and consider alternatives (eg, secure mobile app compatible with personal phones). Fourth, although access to a refrigerator was an eligibility requirement, participants who lived in settings with a shared refrigerator (ie, homeless or domestic violence shelter) had their samples erroneously thrown out. Other participants reported that family members found the samples and, not knowing what they were, threw them out. In these instances, staff issued new supplies. To mitigate this risk, future protocols should include more comprehensive participant education regarding sample storage, as well as considerations for discreet labeling. Finally, the use of cryovials with color-coded caps to visually distinguish waking sample (green) and 30-minute postwaking sample (red) vials may constitute a study limitation for individuals with red-green vision deficits. Future research should consider adopting alternative visual markers (eg, distinct patterns, textual labels, or uniquely shaped indicators) to enhance accessibility and ensure clarity for all participants.

### Future Directions

Future studies that seek to obtain salivary samples among racial and ethnic minority women with histories of abuse should consider the context in which the women are living. In our study, we incorporated safe words and established safety plans with participants during key points of the protocol (ie, sample collection, phone usage, and sample pick-up). In addition, we preprogrammed community resources into the study phone contacts (ie, domestic violence shelters and study contacts) under aliases.

Acknowledging the reality and potential volatility of the context of participants' lives signals support for them and builds trust, which can strengthen protocol adherence. By strengthening adherence, we can advance research that examines the physiological and health repercussions of GBV.

## Acknowledgments

Generative AI was not used in any portion of the manuscript writing.

## Funding

This research was supported by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01HD077891 [JKS, JCC, and KT], the National Institute of Drug Abuse (K01DA031593 [JKS]; T32DA023356 [KT]), the National Institute on Minority Health and Health Disparities (L60MD003701 [JKS]; L60MD011184 [KT]), and the National Institute of Alcohol Abuse and Alcoholism (K01AA025009 [KT]). This research was also supported by the Center for AIDS Research at Johns Hopkins University (P30AI094189 [ANC]) and University of California, San Diego (2P30AI036214-24). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the paper.



XSL•FO
RenderX

## Data Availability

Data access is restricted as the dataset contains sensitive and identifiable information. For this study, an NIH Certificate of Confidentiality was issued, which protects the privacy of research participants involved by prohibiting disclosure of research information to outside requests. Individuals authorized to protect the privacy of research participants are compelled to withhold identifiable characteristics and information. Data requests should be sent to the University of California San Diego Institutional Review Board/Human Research Protections Program (hrpp@ucsd.edu).

## Authors' Contributions

KT contributed to the conceptualization, methodology, investigation, funding acquisition, project administration, resources, formal data analysis, writing (original), and writing (review and editing). MBL contributed to the conceptualization, methodology, investigation, funding acquisition, writing (original), and writing (review and editing). ANC contributed to the conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, and writing. MK contributed to the formal analysis, investigation, software, and writing (review and editing). CJC contributed to the project administration and writing (original). EA contributed to the project administration. EBS contributed to the project administration. JCC contributed to the conceptualization, methodology, investigation, funding acquisition, writing (original), and writing (review and editing). DAG contributed to the conceptualization, methodology, investigation, funding acquisition, writing (original), and writing (review and editing). JKS contributed to the conceptualization, methodology, investigation, funding acquisition, writing (original), and writing (review and editing).

## Conflicts of Interest

DAG is the Chief Scientific and Strategy Advisor at Salimetrics LLC and SalivaBio LLC and these relationships are managed by the policies of the committees on conflict of interest at Johns Hopkins School of Medicine and the University of California at Irvine.

Multimedia Appendix 1
Saliva collection log.
[DOCX File, 31663 KB - i-jmr_v15i1e64150_app1.docx ]

Multimedia Appendix 2
Saliva retrieval form.
[DOCX File, 783 KB - i-jmr_v15i1e64150_app2.docx ]

Multimedia Appendix 3
Saliva decision tree.
[DOCX File, 967 KB - i-jmr_v15i1e64150_app3.docx ]

## References

1. McEwen BS, Stellar E. Stress and the individual. Mechanisms leading to disease. Arch Intern Med 1993 Sep 27;153(18):2093-2101. [Medline: 8379800]
2. Frost DM, Meyer IH. Minority stress theory: application, critique, and continued relevance. Curr Opin Psychol 2023 Jun;51:101579. [doi: 10.1016/j.copsyc.2023.101579] [Medline: 37270877]
3. Calabrese SK, Meyer IH, Overstreet NM, Haile R, Hansen NB. Exploring discrimination and mental health disparities faced by Black sexual minority women using a minority stress framework. Psychol Women Q 2015 Sep 1;39(3):287-304. [doi: 10.1177/0361684314560730] [Medline: 26424904]
4. Sherman ADF, Cimino AN, Balthazar M, et al. Discrimination, sexual violence, depression, post-traumatic stress disorder, and social support among Black women. J Health Care Poor Underserved 2023;34(1):35-57. [doi: 10.1353/hpu.2023.0004] [Medline: 37464480]
5. Mata-Gil S, Sánchez-Cabaco A, Del Moral-Martínez J, Seisdedos-Benito A, Lundberg U. Concentrations of salivary cortisol in victims of intimate partner violence according to the CIRCORT database. Int J Environ Res Public Health 2021 Oct 14;18(20):10819. [doi: 10.3390/ijerph182010819] [Medline: 34682563]
6. Garcia-Moreno C, Jansen H, Ellsberg M, Heise L, Watts CH. Prevalence of intimate partner violence: findings from the WHO multi-country study on women's health and domestic violence. Lancet 2006 Oct 7;368(9543):1260-1269. [doi: 10.1016/S0140-6736(06)69523-8] [Medline: 17027732]
7. Basile KC, Smith SG, Kresnow MJ, Khatiwada S, Leemis RW. The national intimate partner and sexual violence survey: 2016/2017 report on intimate partner violence. : Centers for Disease Control and Prevention; 2022 URL: https://www.cdc.gov/nisvs/documentation/nisvsReportonSexualViolence.pdf [accessed 2026-07-15]



XSL•FO
RenderX

8. Blasco-Ros C, Herbert J, Martinez M. Different profiles of mental and physical health and stress hormone response in women victims of intimate partner violence. Journal of Acute Disease 2014;3(4):303-313. [doi: 10.1016/S2221-6189(14)60066-4]
9. Dillon G, Hussain R, Loxton D, Rahman S. Mental and physical health and intimate partner violence against women: a review of the literature. Int J Family Med 2013;2013(1):313909. [doi: 10.1155/2013/313909] [Medline: 23431441]
10. Lacey KK, Sears KP, Matusko N, Jackson JS. Severe physical violence and Black women's health and well-being. Am J Public Health 2015 Apr;105(4):719-724. [doi: 10.2105/AJPH.2014.301886] [Medline: 24922123]
11. Pico-Alfonso MA, Garcia-Linares MI, Celda-Navarro N, Herbert J, Martinez M. Changes in cortisol and dehydroepiandrosterone in women victims of physical and psychological intimate partner violence. Biol Psychiatry 2004 Aug 15;56(4):233-240. [doi: 10.1016/j.biopsych.2004.06.001] [Medline: 15312810]
12. Seedat S, Stein MB, Kennedy CM, Hauger RL. Plasma cortisol and neuropeptide Y in female victims of intimate partner violence. Psychoneuroendocrinology 2003 Aug;28(6):796-808. [doi: 10.1016/s0306-4530(02)00086-0] [Medline: 12812865]
13. Alhalal E, Falatah R. Intimate partner violence and hair cortisol concentration: a biomarker for HPA axis function. Psychoneuroendocrinology 2020 Dec;122:104897. [doi: 10.1016/j.psyneuen.2020.104897] [Medline: 33068953]
14. Kim HK, Tiberio SS, Capaldi DM, Shortt JW, Squires EC, Snodgrass JJ. Intimate partner violence and diurnal cortisol patterns in couples. Psychoneuroendocrinology 2015 Jan;51:35-46. [doi: 10.1016/j.psyneuen.2014.09.013] [Medline: 25286224]
15. Feinberg ME, Jones DE, Granger DA, Bontempo D. Relation of intimate partner violence to salivary cortisol among couples expecting a first child. Aggress Behav 2011;37(6):492-502. [doi: 10.1002/ab.20406] [Medline: 21830223]
16. Segerstrom SC, Boggero IA, Smith GT, Sephton SE. Variability and reliability of diurnal cortisol in younger and older adults: implications for design decisions. Psychoneuroendocrinology 2014 Nov;49:299-309. [doi: 10.1016/j.psyneuen.2014.07.022] [Medline: 25137484]
17. Golden SH, Sánchez BN, DeSantis AS, et al. Salivary cortisol protocol adherence and reliability by socio-demographic features: the multi-ethnic study of atherosclerosis. Psychoneuroendocrinology 2014 May;43:30-40. [doi: 10.1016/j.psyneuen.2014.01.025]
18. Chida Y, Steptoe A. Cortisol awakening response and psychosocial factors: a systematic review and meta-analysis. Biol Psychol 2009 Mar;80(3):265-278. [doi: 10.1016/j.biopsycho.2008.10.004] [Medline: 19022335]
19. Stalder T, Lupien SJ, Kudielka BM, et al. Evaluation and update of the expert consensus guidelines for the assessment of the cortisol awakening response (CAR). Psychoneuroendocrinology 2022 Dec;146:105946. [doi: 10.1016/j.psyneuen.2022.105946] [Medline: 36252387]
20. Dockray S, Bhattacharyya MR, Molloy GJ, Steptoe A. The cortisol awakening response in relation to objective and subjective measures of waking in the morning. Psychoneuroendocrinology 2008 Jan;33(1):77-82. [doi: 10.1016/j.psyneuen.2007.10.001] [Medline: 17996380]
21. Broderick JE, Arnold D, Kudielka BM, Kirschbaum C. Salivary cortisol sampling compliance: comparison of patients and healthy volunteers. Psychoneuroendocrinology 2004 Jun;29(5):636-650. [doi: 10.1016/S0306-4530(03)00093-3] [Medline: 15041086]
22. Smyth N, Thorn L, Hucklebridge F, Evans P, Clow A. Detailed time course of the cortisol awakening response in healthy participants. Psychoneuroendocrinology 2015 Dec;62:200-203. [doi: 10.1016/j.psyneuen.2015.08.011] [Medline: 26318630]
23. Miller F, Miller P. Transgenerational trauma and trust restoration. AMA J Ethics 2021 Jun 1;23(6):E480-E486. [doi: 10.1001/amajethics.2021.480] [Medline: 34212849]
24. Randolph SD, Golin C, Welgus H, Lightfoot AF, Harding CJ, Riggins LF. How perceived structural racism and discrimination and medical mistrust in the health system influences participation in HIV health services for Black women living in the United States South: a qualitative, descriptive study. J Assoc Nurses AIDS Care 2020;31(5):598-605. [doi: 10.1097/JNC.0000000000000189] [Medline: 32868634]
25. Cox K. Black Americans and mistrust of the US health care system and medical research. : Pew Research Center; 2024 URL: https://www.pewresearch.org/race-and-ethnicity/2024/06/15/black-americans-and-mistrust-of-the-u-s-health-care-system-and-medical-research/ [accessed 2026-05-29]
26. Benkert R, Cuevas A, Thompson HS, Dove-Meadows E, Knuckles D. Ubiquitous yet unclear: a systematic review of medical mistrust. Behav Med 2019;45(2):86-101. [doi: 10.1080/08964289.2019.1588220] [Medline: 31343961]
27. Correlates of perceived discrimination in healthcare amongst Black women. GSJP 2022;18:18. [doi: 10.52214/gsjp.v18i.10927]
28. Souleymanov R, Akinyele-Akanbi B, Njeze C, et al. Navigating ethics in HIV data and biomaterial management within Black, African, and Caribbean communities in Canada. BMC Med Ethics 2025 Jan 16;26(1):5. [doi: 10.1186/s12910-025-01161-0] [Medline: 39815313]
29. Sheppard VB, Mays D, LaVeist T, Tercyak KP. Medical mistrust influences black women's level of engagement in BRCA 1/2 genetic counseling and testing. J Natl Med Assoc 2013;105(1):17-22. [doi: 10.1016/s0027-9684(15)30081-x] [Medline: 23862292]




XSL•FO
RenderX

30. Tekeste M, Hull S, Dovidio JF, et al. Differences in medical mistrust between Black and White women: implications for patient-provider communication about PrEP. AIDS Behav 2019 Jul;23(7):1737-1748. [doi: 10.1007/s10461-018-2283-2] [Medline: 30264207]
31. Grove M, et al. The legacy effect: understanding how segregation and environmental injustice unfold over time in baltimore. In: Social Justice and the City: Routledge; 2020:224-237.
32. Henrietta Lacks: science must right a historical wrong. Nature 2020 Sep;585(7823):7-7. [doi: 10.1038/d41586-020-02494-z] [Medline: 32873976]
33. Pratt-Harris NC, Sinclair MM, Bragg CB, et al. Police-involved homicide of unarmed Black males: observations by Black scholars in the midst of the April 2015 Baltimore uprising. J Hum Behav Soc Environ 2016 May 18;26(3-4):377-389. [doi: 10.1080/10911359.2015.1132853]
34. Saliva Collection and Handling Advice. : Salimetrics LLC; 2011 URL: https://www.lubio.ch/assets/Salimetrics_Saliva_Collection_Handbook.pdf [accessed 2026-07-15]
35. Nasser A, Ozenne B, Høgsted ES, Jensen PS, Frokjaer VG. Reliability of three versus five saliva sampling times for assessing the cortisol awakening response. Psychoneuroendocrinology 2023 Jan;147:105950. [doi: 10.1016/j.psyneuen.2022.105950] [Medline: 36272363]
36. Granger DA, Johnson SB, Szanton SL, Out D, Schumann LL. Incorporating salivary biomarkers into nursing research: an overview and review of best practices. Biol Res Nurs 2012 Oct;14(4):347-356. [doi: 10.1177/1099800412443892] [Medline: 22593229]
37. StataCorp LLC | Stata. URL: https://www.stata.com/company/ [accessed 2026-07-15]

## Abbreviations

**CAR:** cortisol awakening response
**ESSENCE:** Examining Stress, Sexual Experiences, and Neighborhood Correlates of HIV Risk among Black Women
**GBV:** gender-based violence
**HPA:** hypothalamic-pituitary-adrenal
**IPV:** intimate partner violence
**STD:** sexually transmitted disease

*Edited by M Balcarras; submitted 19.Jul.2024; peer-reviewed by CA Clair, L Marquez-Magana, RSG Mahmoud, T Theorell; revised version received 06.Apr.2026; accepted 08.Apr.2026; published 03.Sep.2026.*

*Please cite as:*
*Tsuyuki K, Lucea MB, Cimino AN, Killian M, Catabay CJ, Abrams E, Baffour-Siaw E, Campbell JC, Granger DA, Stockman JK Repeat At-Home Saliva Collection to Measure Cortisol Awakening Response (CAR): Protocol Feasibility Study Among Black Women With Histories of Abuse*
*Interact J Med Res 2026;15:e64150*
*URL: https://www.i-jmr.org/2026/1/e64150*
*doi:10.2196/64150*

© Kiyomi Tsuyuki, Marguerite B Lucea, Andrea N Cimino, Michael Killian, Christina J Catabay, Engle Abrams, Emmanuel Baffour-Siaw, Jacquelyn C Campbell, Douglas A Granger, Jamila K Stockman. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 3.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Effectiveness and Lessons Learned From an Occupational E-Mental Health Intervention for Enhancing Workplace Mental Health: The EMPOWER Cluster Randomized Controlled Trial

Carlota de Miquel[1,2,3], PhD; Christina M Van der Feltz-Cornelis[4,5], MD, PhD; Leona Hakkaart-van Roijen[6], PhD; Dorota Merecz-Kot[7], PhD; Marjo Sinokki[8,9], PhD; Jordi Rodeiro-Boliart[1], MSc; Jennifer Sweetman[4], PhD; Kaja Staszewska[7], MA; Ellen Vorstenbosch[10], MSc; Daniele Porricelli[1], MSc; Stijn Peeters[6], MSc; José Luis Ayuso-Mateos[2,11,12], MD, PhD; Luis Salvador-Carulla[13,14], MD, PhD; Sue Lukersmith[13], PhD; Oriol Borrega[15], MSc; Carla Sabariego[10,16,17], PhD; Christophe Vanroelen[18], PhD; Alberto Raggi[19], PhD; Diletta Porcheddu[20], PhD; Josep Maria Haro[1,2,3,21], MD, PhD; Beatriz Olaya[1,2,22], PhD

[1]Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain

[10]Swiss Paraplegic Research, Nottwil, Lucerne, Switzerland

[11]Department of Psychiatry, Hospital Universitario de La Princesa, Instituto de Investigación Sanitaria del Hospital Universitario de La Princesa, Madrid, Spain

[12]Department of Psychiatry, Universidad Autónoma de Madrid, Madrid, Spain

[13]Health Research Institute, University of Canberra, Canberra, Australia

[14]Healthcare Information Systems (CTS553), Universidad de Cádiz, Cadiz, Spain

[15]Òmada Interactiva, SLL, Barcelona, Spain

[16]Faculty of Health Sciences and Medicine, University of Lucerne, Lucerne, Switzerland

[17]Center for Rehabilitation in Global Health Systems, University of Lucerne, Lucerne, Switzerland

[18]Brussels Institute for Social and Population Studies, Vrije Universiteit Brussel, Brussels, Belgium

[19]Neurology, Public Health and Disability Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy

[2]CIBERSAM, Centro de Investigación Biomédica en Red de Salud Mental, Instituto de Salud Carlos III, Madrid, Spain

[20]Fondazione ADAPT, Milan, Italy

[21]Research, Innovation and Teaching Unit, Parc Sanitari Sant Joan de Déu, Sant Boi de Llobregat, Spain

[22]Department of Clinical and Health Psychology, Universitat Autònoma de Barcelona, Faculty of Psychology, Building B, Campus de la UAB, Bellaterra (Cerdanyola del Vallès), Spain

[3]Department of Medicine, Universitat de Barcelona, Barcelona, Spain

[4]Department of Health Sciences, University of York, York, United Kingdom

[5]Institute of Health Informatics, University College London, London, United Kingdom

[6]Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

[7]Institute of Psychology, University of Łódź, Łódź, Poland

[8]Länsirannikon Työterveys Oy, Turku, Finland

[9]Unit of Public Health, University of Turku, Turku, Finland

**Corresponding Author:**
Beatriz Olaya, PhD
Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Spain

## Abstract

**Background:** Occupational e-mental health (OeMH) interventions emerged as a promising solution to prevent common mental health problems and enhance well-being and work performance. However, they must be subject to robust and reliable assessments for effectiveness.

**Methods:** A multimodal e-mental health intervention (EMPOWER [The European Platform to Promote Wellbeing and Health in the Workplace]) delivered over 7 weeks was developed and evaluated through a cluster randomized controlled trial conducted mainly in small to medium enterprises and public agencies from Spain (n=127), Finland (n=141), Poland (n=51), and the United Kingdom (n=389) between February 2022 and May 2024 (recruitment finalized in September 2023 and follow-up completed in May 2024). Inclusion criteria were being 18+ years, having a smartphone, sufficient language knowledge, and agreeing to participate. Clusters (companies or departments) were randomized to intervention or control conditions. The primary outcome



XSL•FO
RenderX

was presenteeism, and secondary outcomes were depression and anxiety symptoms, etc, all measured at baseline, postintervention, and in 21 weeks after program completion. The analysis was performed as an intention-to-treat approach using adjusted linear mixed models and as per protocol analysis comparing outcomes by level of engagement.

**Results:** A total of 347 participants were allocated to the intervention group and 361 to the control group. In the overall sample, the intention-to-treat analysis detected no statistically significant short-term (7 wk) or long-term (21 wk postintervention) effects of the EMPOWER intervention on presenteeism (postintervention β=2.186; 95% CI −2.424 to 6.796, follow-up β=1.294; 95% CI −3.608 to 6.396) and on other secondary outcomes such as depressive symptoms (postintervention β=−0.052, 95% CI −1.02 to 0.905, follow-up β=0.202, 95% CI −0.840 to 1.245), anxiety symptoms (postintervention β=−0.328, 95% CI −1.168 to 0.512, follow-up β=0.375, 95% CI −0.537 to 1.287), or general stress level (postintervention β=0.385, 95% CI −0.195 to 0.965, follow-up β=0.123, 95% CI −0.502 to 0.749). Subgroup analyses yielded several notable results, with significant differences between age groups, gender, and psychological symptoms at baseline. The per-protocol analysis showed no significant differences between participants who actively engaged with the intervention (119/347, 34%) and those who did not. Implementation challenges were related to technical problems, the complexity of this study's design, external factors, co-design strategy, and organizational barriers, which led to a smaller sample size, high attrition rates, and low adherence.

**Conclusions:** Our study provides evidence from a large cluster randomized controlled trial evaluating an OeMH intervention implemented in workplace settings, including small to medium enterprises and public agencies in Europe. Although no overall effectiveness was observed, this study contributes important methodological and implementation insights, highlighting the challenges of evaluating OeMH interventions. These findings suggest that future interventions should prioritize feasibility testing, organizational readiness, user engagement, and more targeted and pragmatic evaluation approaches to enhance real-world impact.

**Trial Registration:** ClinicalTrial.gov NCT04907604; https://clinicaltrials.gov/study/NCT04907604

**International Registered Report Identifier (IRRID):** RR2-10.1177/20552076221131145

***(Interact J Med Res 2026;15:e66041)*** doi:10.2196/66041

**KEYWORDS**

occupational e-mental health intervention; workplace; RCT; effectiveness; implementation challenges; e-mental health; mental health; digital health intervention; digital health interventions; depression; anxiety; stress; well-being; well being; insomnia; randomized controlled trial

## Introduction

Common mental health conditions, such as depression, anxiety, or other stress-related disorders, constitute approximately 40% of work-related illness cases, and these conditions are mostly manageable and avoidable with appropriate treatment and prevention [1,2]. On a global scale, around 12 billion workdays are forfeited annually due to the impact of depression and anxiety, resulting in an economic cost of approximately US $1 trillion per year attributed to reduced productivity [3]. In Europe, productivity loss due to mental health conditions has an estimated annual cost of €400 billion (≈US $460.4 billion) [4,5]. This productivity loss is a consequence of working with limitations due to an illness (ie, presenteeism) and absence from work due to the illness (ie, absenteeism) [6].

Occupational mental health interventions have been increasingly delivered through digital platforms, described as being more flexible, anonymous, and convenient [7]. This has the potential to increase the accessibility of evidence-based interventions and to reduce costs compared to face-to-face interventions [8]. There is evidence for moderate effects of occupational e-mental health (OeMH) interventions to reduce stress, insomnia, and burnout, and small treatment effects on depression, anxiety, well-being, and mindfulness [9-11]. Results regarding work-related outcomes, such as presenteeism and absenteeism, are scarce and inconclusive [10,12]. In general, there is a high heterogeneity between studies on OeMH interventions in terms of the type of intervention delivered, outcomes, and the quality of their methodology [13,14]. Furthermore, these mixed results underscore the complexity of assessing the outcomes of digital interventions and indicate that organizational culture, intervention design, and implementation serve as critical mediating factors.

In fact, the implementation of e-mental health interventions in the workplace has some important barriers. Challenges include adapting to new and changing regulatory frameworks, ensuring compatibility with existing systems, addressing employee privacy and security concerns, the rapid change of the digital ecosystem, and managing costs [15]. Additionally, maintaining treatment adherence and preventing attrition are also recognized as challenges. e-Mental health interventions have usually been associated with high attrition rates [8]. A recent scoping review indicates that potential successful implementation strategies for OeMH interventions include providing incentives to participants, sending reminders, providing support to users during the intervention, conducting educational meetings, and the involvement of senior management [16]. The scoping review also concludes that there is an urgent need to enhance the reporting of implementation strategies and to integrate common implementation frameworks with more technology-focused ones to comprehensively address the complexities of e-health implementations [16].

Several factors can potentially affect the effectiveness of OeMH interventions, such as the mode of delivery, the content, the approach (eg, tailored and universal), duration of intervention, the use of interventions in combination, and the characteristics



XSL•FO
RenderX

of the target population [17]. Digital interventions that are tailored to users' needs tend to foster higher engagement [17]. On the other hand, stigma against individuals with mental illness can deter participation in such interventions [18]. Finally, the cultural adaptation of workplace interventions is a critical factor for successful implementation, with more favorable outcomes for linguistically and culturally adapted interventions [19] and those that consider organizational cultural elements [20]. Culturally adapted e-mental health interventions seem to have a moderately higher effect in reducing depressive and anxiety symptoms when compared to interventions that are not culturally adapted [19].

While occupational mental health interventions delivered via digital platforms have emerged as promising tools for addressing workplace mental health issues, these interventions often lack contextual adaptation to workplace-specific needs and exhibit high variability in outcomes due to differences in design, content, and target populations. A systematic review by Moe-Byrne et al [12] highlighted this variability in the effectiveness of digital workplace mental health interventions and emphasized the need for context-sensitive, tailored approaches. Additionally, although digital mental health interventions are growing, many existing programs do not sufficiently consider organizational cultural factors or provide tailored support for small to medium enterprises (SMEs) [21].

The EMPOWER [The European Platform to Promote Wellbeing and Health in the Workplace] platform is a multimodal eHealth intervention delivered through a website and an application that was designed to address mental health issues in the workplace and enhance employees' well-being [22]. It follows a 3-tiered intervention structure: universal (primary) prevention; targeted (secondary) prevention, and tertiary prevention, focusing primarily on preventing mild mental health problems and promoting well-being in the workplace rather than treating mental disorders. It was specifically designed for SMEs and public agencies (for a comprehensive view of the design and rationale of the current study, please see references [22,23]), filling a critical gap identified in recent literature. EMPOWER integrates organizational readiness assessments and user-centered design principles to enhance feasibility and effectiveness, aiming to improve outcomes such as presenteeism, anxiety, depression, and well-being. We conducted a cluster randomized controlled trial (RCT) with employees from SMEs and public agencies in Finland, Poland, Spain, and the United Kingdom to evaluate the effectiveness of EMPOWER on various mental health and work-related outcomes, both immediately postintervention and at follow-up, and explored several factors that could impact the effectiveness through subgroup analyses. The rationale behind using a cluster randomized design was the organizational nature of the intervention and the high risk of contamination between employees within the same workplace. The primary and secondary outcomes were assessed at the individual employee level, while randomization and intervention allocation were performed at the organizational (cluster) level. Additionally, we discuss the challenges encountered during the implementation process and interpret our results in light of these difficulties.

## Methods

### Study Design

This study was conducted as a multicountry cluster RCT with 2 parallel arms (intervention and wait-list control), following an initial stepped-wedge trial design in Spain, Finland, Poland, and the United Kingdom. Clusters (ie, companies or departments) were randomly assigned to A, B, and C groups, differing in the time when participants had access to the intervention. Participants from the 3 groups were invited to download the application at the same time, that is, at the beginning of the RCT (T0). Participants in group A had access to the intervention after completing the baseline questionnaires (T0), participants in group B had access after completing 2 rounds of questionnaires (after T1), and participants in group C had access after completing 3 rounds of questionnaires (after T2). A total of 2 additional rounds of questionnaires were delivered to the participants as follow-up measures (T3 and T4). As such, data were collected at the individual level at 5 time points [22]. Each data collection wave took place with a span of approximately 7 weeks of difference. The first study protocols were approved by the ethics committees of Fundació Sant Joan de Déu (PIC-39‐20), Turku University Hospital (PIC-993966082), University of York (HSRGC250321), and Institute of Occupational Medicine, University of Lodz (9/2020).

Periodic reports and meetings were held among the consortium partners to monitor the progress of the fieldwork. We faced several barriers and challenges throughout the trial period, particularly during the first months (February 2022 to September 2022). During this time, various contingency measures were implemented to increase participation, including the inclusion of new companies, improvements in local dissemination activities, and sending reminders to workers. In Table S1 in Multimedia Appendix 1, we present the contextual barriers (at the country, company, and individual levels) as well as the facilitators identified throughout the trial.

In September 2022, the number of recruited participants was still lower than expected, despite the contingency actions that we took (see section Sample Size Calculation; 461 had consented to participate, 295 had completed baseline questionnaires, 100 had answered T1, 45 had answered T2, and 14 had answered T3). Feedback from participants, collected through qualitative interviews, revealed that one of the main reasons for not completing the planned assessments was the length of the questionnaires and the high number of assessment points. After careful consideration, the consortium agreed to modify the design from a stepped-wedge to a cluster RCT with 2 parallel arms: an intervention arm and a wait-list control arm. Following this change, each newly recruited cluster (company or department) was randomized to either intervention or control, regardless of size, type of workers, or sector. We also reduced the number of assessment points from 5 to 3 in all groups: T0 (baseline), T1 (postintervention, 7 weeks), and T2 (follow-up, 21 weeks). In addition, the number of questionnaires at T1 was reduced to minimize participant burden.

Participants recruited before September 2022, and therefore under the initial stepped-wedge design, were reallocated to



XSL•FO
RenderX

either the intervention or control group depending on their study status. Specifically, participants from group A, who had already completed the intervention between T0 and T1, were directly assigned to the intervention group. Participants from group B were reassigned to the intervention group, with their T1 data treated as preintervention, T2 as postintervention, and their follow-up postponed by 7 weeks to ensure comparability with group A. Finally, participants in group C continued their original assessment schedule, with the exception that they did not gain access to the intervention until the follow-up assessment (21 weeks).

This study's design modification was approved by Fundació Sant Joan de Déu in October 2022. The trial was registered at ClinicalTrial.gov (NCT04907604). The trial ended as planned after completion of follow-up assessments.

This cluster RCT was reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) 2010 statement [24], the CONSORT extension for cluster randomized trials [25], and the CONSORT extension for abstracts [26] (Checklist 1).

### Recruitment Process and Randomization

Before the recruitment of participants, we approached several SMEs and public agencies from Finland, Poland, Spain, and the United Kingdom to participate in this study. Large companies were also considered for the RCT. Clusters were eligible if they were SMEs, public agencies, or large organizations willing to participate in this study and able to provide access to their employees for the duration of the trial. There was no limitation on the size or economic sector. Participant recruitment took place between February 25, 2022, and September 30, 2023; data collection lasted until May 30, 2024. New companies or departments could be included during the course of this study.

Different strategies were used by the local research teams to recruit companies (such as press release and media, employers' organizations, and face-to-face and online meetings). A formal agreement was signed between the EMPOWER consortium and the employer before recruiting their employees. In a first round of randomization (January 2022) and under the stepped-wedge trial design, clusters (ie, SMEs or departments) were randomly allocated into group A, B, or C, considering strata to reduce imbalance issues [27]. This was done separately in each country, considering, when possible, the cluster size, type of workers (blue vs white collar), and type of company (SME, public agency, and large). A second round of randomization was repeated in April 2022. We used the runiform module from Stata (StataCorp LLC) to randomize newly recruited companies or departments. After the design modification with 2 parallel groups (from September 2022), newly recruited clusters were randomly allocated to the intervention or control group by using randomly generated numbers in Excel (Microsoft Corp). In this case, variables such as cluster size, type of workers, and type of company were not taken into account when randomizing the companies. Randomization was performed by this study's coordinator. Moreover, due to the nature of the digital intervention, neither participants nor researchers were blinded to group allocation. Allocation concealment was not feasible due to the cluster-level nature of the intervention and the pragmatic implementation of the trial.

A total of 25 clusters were assigned to the intervention group and 19 to the control group. Initially, email invitations to the employees using the EMPOWER email address were sent, which provided a nontransferable, one-use link for downloading the application and a company-specific code to access it. This allowed us to have control over participants, but at the same time, did not allow participants to share the link with other workmates and increased the chance of technical problems. Thus, in April 2022, we created QR codes for each country to download the application, together with a company code to sign in. All procedures followed the General Data Protection Regulation [28].

Participant inclusion criteria were (1) being aged 18 years or older, (2) having a mobile phone with internet access, (3) having sufficient knowledge of the local language, and (4) providing informed consent.

All participants were invited to download the application and provide informed consent, which was integrated into the application. After providing consent, participants were asked to answer the baseline assessment questions. Employees of localities assigned to the intervention group were able to access and use the intervention part of the application directly after completing the baseline questionnaires; application access lasted for a duration of 7 weeks. All participants were asked to complete additional rounds of questionnaires 7 weeks after baseline assessment completion (postintervention) and 21 weeks after postintervention assessment completion (follow-up). Once the 3 rounds of questionnaires were completed, participants in both groups (intervention and control groups) had full access to the intervention part of the application. Participants with severe mental health conditions (including people reporting suicidal thoughts) requiring immediate treatment were advised to seek professional help and were warned that the EMPOWER application was not an intervention; however, they were not excluded from this study.

The assessment protocols were delivered through the application with no time limit for participants to complete the questionnaires. Participants received automatic reminder emails and/or pop-up messages from the application to notify them when an assessment protocol was available for them to complete. To increase retention rates, weekly reports were used to follow up on the number of responses, and the local researchers actively sent personalized email reminders to participants to encourage them to answer the questionnaires. Local researchers in each country had a coordination handbook that covered topics such as study design, data management and security, ethical aspects, responsibilities and good practices, solving problems during the fieldwork, and the incidental findings policy. No specific procedures were implemented to monitor adverse events, as the intervention was low-intensity and self-guided. Any potential unintended effects were assessed descriptively. Participants were provided with the contact details of the local research teams.



XSL•FO
RenderX

### Sample Size Calculation

A cautionary approach was adopted to detect small changes and effect sizes and was based on a review that assessed the impact of mental health programs on presenteeism in the workplace [29]. The calculation also accounted for potential loss to early and late follow-ups (25%). The sample size calculation was based on detecting changes in presenteeism, based on a review of the impact of mental health programs on presenteeism in the workplace [30]. Considering an expected effect size of $d$=0.30, a bilateral type I error of 0.05, and a power of 80%, a total sample of 729 participants was required. However, due to the implementation of multilevel analyses resulting from cluster randomization, the sample size needed to be adjusted by a factor of 1.2, leading to a final sample size of 874 participants [29].

### Ethical Considerations

This study was conducted in compliance with ethical guidelines for research involving human participants. Ethical approval was obtained from the relevant ethics committees in each participating country: Clinical Research Ethics Committee of Fundació Sant Joan de Déu (PIC-39-20), Ethics Committee of the Hospital District of Southwest Finland (PIC-993966082), Health Sciences Research Governance Committee, University of York (HSRGC250321), and the Institute of Occupational Medicine and the Bioethics Committee of the Nofer Institute of Occupational Medicine (Łódź, Poland; 9/2020). The trial was registered at ClinicalTrials.gov (NCT04907604). This study's design modifications made in response to recruitment challenges were approved by Fundació Sant Joan de Déu in October 2022.

All participants provided informed consent before enrollment in this study. Consent was integrated into the EMPOWER application, where participants were required to review study information and explicitly agree to participate before accessing the baseline questionnaire. Participants had the right to withdraw from this study at any time without consequence. Participants who completed all 3 rounds of questionnaires (baseline, post intervention at 7 wk, and follow-up at 21 wk) received a €20 (US $22.94) voucher as compensation for their time and effort in this study.

All data collected in this study were anonymized and stored securely in compliance with the General Data Protection Regulation. Each participant was assigned a unique identification code to ensure confidentiality. Identifiable personal data was not collected, and all responses were encrypted during data transmission and storage. Employers had access only to aggregated, anonymized workplace risk assessments, ensuring that individual participant data remained confidential.

### Intervention

The EMPOWER digital intervention has been previously described in detail [23]. The intervention included components delivered at both the individual employee level (self-guided application modules) and the organizational level (employer web portal and aggregated psychosocial risk feedback). The website is public and offers a company-level mental health awareness campaign that assists both employees and employers in addressing work-related stress and mental health challenges. The application is available for use by individual employees and contains triage assessment tools and an algorithm to tailor content modules based on their symptomatology and work functioning. The application follows a self-guided approach, and the content is based on cognitive behavioral therapy (CBT) techniques, including psychoeducational material, tracking your mood, self-guided goal setting, problem solving, relaxation, and breathing exercises. The different modules of the application were designed to increase awareness about mental health, reduce stress, anxiety, and other psychological symptoms, screen for psychosocial risk factors and recommendations, and include a return-to-work module. From an organizational level, employers could access a restricted web portal that provided an anonymous brief assessment of psychosocial risk factors in the workplace as reported by employees, accompanied by tailored recommendations for employers. The readiness of the website and application was evaluated using the technology readiness level adapted for implementation sciences [31] as part of the impact assessment, which is described elsewhere [32]. During the RCT, we collected feedback from employees, employers, and local researchers as part of a qualitative assessment and impact analysis, described elsewhere [32].

### Outcome Measures

All primary and secondary outcomes were assessed at the individual employee level. A summary of the outcome measures used in this study can be found in Table 1. Levels of anxiety and depression were categorized using cutoff scores from the GAD-7 (Generalized Anxiety Disorder Questionnaire) and PHQ-9 (Patient Health Questionnaire-9) scales, where scores greater than 5 are considered as having at least mild anxiety symptoms [33] or mild depressive symptoms [34].



XSL•FO
RenderX

**Table .** Variables and measures for this study.

| Outcome | Measure | Description |
|---|---|---|
| Primary outcome measure | | |
| Presenteeism | iMTA[a] Productivity Cost Questionnaire (iPCQ) [35] | Scores start at 0 and measure total hours lost due to being unproductive at work |
| Secondary outcome measures | | |
| Depression | Patient Health Questionnaire-9 (PHQ-9) [34] | Its total score ranges from 0 to 27, ranging from no depression to severe depression. |
| Anxiety | Generalized Anxiety Disorder Questionnaire (GAD-7) [33] | This questionnaire ranges from 0 to 21, meaning from no anxiety to severe anxiety. |
| Insomnia | Insomnia Severity Index (ISI) [36] | The ISI ranges from 0 to 12, where 0 indicates no presence of insomnia, and 12 indicates a severe presence of insomnia. |
| Work stress | Mini-Psychosocial Stressors at Work Scale (Mini-PSWS) [37] | The total score ranges from 0 (no stress) to 48 (maximum stress) |
| General stress | Perceived Stress Scale (PSS-4) [38] | The total score of the PSS-4 ranges from 0 to 16, where 0 indicates no stress and 16 indicates severe stress. |
| Well-being | World Health Organization Well-Being Index (WHO-5) [39] | It scores from 0 to 100, where higher scores indicate better well-being, and a clinically significant change was defined as a difference of 10 points or more. |
| Momentaneous well-being | Visual analog scale from 0 to 10 | Higher scores represent better well-being |
| Mental Health Quality of Life | Mental Health Quality of Life Questionnaire (MHQoL) [40] | Ranging from 0 to 21, where 21 represents the best possible mental health and quality of life. |
| Physical activity | International Physical Activity Questionnaire (IPAQ) [41] | Reported as total METs[b], with higher values indicating greater activity levels. |
| Sitting | International Physical Activity Questionnaire (IPAQ) [41] | Was recorded in minutes per day spent sitting |
| Somatization | Patient Health Questionnaire (PHQ-15) [42] | Scores from 0 to 30, where 30 indicates the worst health |
| Absenteeism | iMTA Productivity Cost Questionnaire (iPCQ) [35] | Scores start at 0 and measure total hours lost due to being absent from work |
| Unpaid labor | iMTA Productivity Cost Questionnaire (iPCQ) [35] | It calculates the hours lost due to needing assistance with tasks |

[a]iMTA: Institute for Medical Technology Assessment.

[b]MET: metabolic equivalent of task.

## Level of Engagement (Active vs Inactive Application Users)

In our analysis of the intervention group, participants were categorized as either active or inactive application users based on several key criteria. For each participant, we evaluated whether they earned more than 5 badges, engaged with any content by reading pages, began working on any problems, consistently tracked their mood for more than a week (7 d), or acquired any badge through application interaction. Participants were assigned a score of 1 if a condition was met and 0 if it was not. We then summed these scores for each participant. Those who met none or only one of these conditions were classified as inactive, while participants who met more than 1 condition were classified as active.

## Statistical Analysis

For descriptive analysis, we used mean and SD, and frequencies and percentages (%) for continuous and categorical variables, respectively. Differences between groups were calculated using chi-squared and *t* tests. We also calculated a correlation matrix between the main outcomes at baseline to assess the relationships between those variables (see Table S2 in Multimedia Appendix 1).

As a first step, we followed an intention-to-treat approach to maintain the comparability between participants in the intervention and the control groups and avoid selection bias [43]. We calculated linear mixed models (LMMs) to separately analyze the primary and secondary outcomes. The independent variables were the participant's group (intervention vs control) and the intervention time (baseline, postintervention, and follow-up). Each model was adjusted for the covariates gender,



XSL•FO
RenderX

age, and country to account for potential confounding factors. Additional analyses were also performed, adding the same baseline measure as a covariate, to control for potential group differences at baseline. For each outcome, a *P* value and CIs are provided. LMMs were calculated for the whole sample. To investigate potential moderators for the effectiveness of the intervention on several outcomes, we conducted subgroup analyses where LMMs were estimated separately for different subgroups: gender, age groups (18 - 35, 36 - 49, and 50+ years), low and high levels of anxiety, depressive and somatization symptoms (using the clinical cutoff point of 5), low and high levels of work stress (using median as cutoff). For each subgroup, LMMs were adjusted for sex, age, and country to account for potential confounding factors (these covariates were held constant in each of the models; information about the coefficients for each covariate in the LMM models is available upon request).

Statistics that compared participants who dropped out with those who did not were also described with mean (SD) or n (%), depending on whether the outcomes were numeric or categorical. Finally, we conducted a per-protocol analysis as a secondary analysis, in which we conducted LMMs differentiating between active and inactive participants (see section Level of Engagement (Active vs Inactive Application Users)), comparing them to the control condition and between each other. As such, the per-protocol analyses in our study did not handle missing data. Per-protocol analysis estimates the effect of receiving the treatment (efficacy). Missing data were handled by using LMMs, which allow inclusion of participants with incomplete data under the missing at random assumption.

## Results

### Descriptive Analyses

A total of 877 employees accepted the informed consent, and 708 answered the baseline questionnaires. A total of 44 clusters were randomized, with 25 allocated to the intervention group and 19 to the control group. No clusters were lost or excluded after randomization, and all of them were included in the primary analyses. A total of 333 participants completed postintervention questionnaires and 267 follow-up questionnaires (Figure 1). Table 2 shows the descriptive statistics for the total sample and by intervention group, 78% (544/697) were females, and the average age was 45.05 (SD 10.50) years, 18% (127/708) of participants were from Spain, 7% (51/708) from Poland, 55% (389/708) from the United Kingdom, and 20% (141/708) from Finland. Most participants were working in public agencies (584/685, 85%) and were white-collar workers (663/696, 95%). The intervention and control groups significantly differed for sex, country, company, and workers' type. All participants who completed baseline assessments were included in the intention-to-treat analyses for the primary outcome.



XSL•FO
RenderX

**Figure 1.** Participant flow diagram.



XSL•FO
RenderX

**Table .** Sociodemographic sample descriptive statistics for the total sample and by intervention group.

| Variable | Total sample | Control group | Intervention group | Differences | |
|---|---|---|---|---|---|
| | | | | Test (*df*) | *P* value |
| Gender, n (%) | | | | 8.33 (2)[a] | .02 |
| Female | 544 (78.05) | 292 (82) | 252 (73.47) | | |
| Age, mean (SD) | 44.20 (10.52) | 44.17 (10.03) | 44.23 (11.01) | 0.126 (688.8)[b] | .88 |
| Country, n (%) | | | | 22.538 (3)[a] | .001 |
| Spain | 127 (17.94) | 48 (13.30) | 79 (22.77) | | |
| Poland | 51 (7.20) | 17 (4.71) | 34 (9.80) | | |
| United Kingdom | 389 (54.94) | 225 (62.33) | 164 (47.26) | | |
| Finland | 141 (20) | 71 (20) | 70 (20) | | |
| Company type, n (%) | | | | 39.017 (2)[a] | .001 |
| Public agency | 584 (85.26) | 327 (90.58) | 257 (79.32) | | |
| SME[c] | 90 (13.14) | 23 (6.37) | 67 (20.68) | | |
| Large company | 11 (1.61) | 11 (3.05) | 0 (0) | | |
| Blue or white collar, n (%) | | | | 13.338 (1)[a] | .001 |
| White collar | 663 (95.26) | 347 (98.30) | 316 (92.13) | | |

[a]Chi-square test.

[b]*t* test.

[c]SME: small to medium enterprise.

In general, the correlation matrix showed high correlation coefficients between most outcomes at baseline (except for the variables sitting and absenteeism, see Table S2 in Multimedia Appendix 1). Table 3 shows the mean values for the outcomes at preintervention, generally, and divided by intervention and control groups. At preintervention (T0), the intervention group had significantly lower levels of depressive symptoms ($t$=2.031, $P$=.043), anxiety symptoms ($t$=2.169, $P$=.03), and general stress ($t$=3.104, $P$=.002), and higher levels of well-being ($t$=−3.412, $P$=.001), mental health quality of life ($t$=−2.166, $P$=.031), and momentaneous well-being ($t$=−2.438, $P$=.015), as compared with the control group. No statistically significant differences between groups were found for insomnia, work stress, physical activity, sitting, somatic health, absenteeism, presenteeism, and unpaid labor (all $P$>.05).



XSL•FO
RenderX

**Table .** Baseline means (SDs) and group comparisons for primary and secondary outcome variables.

| Variable | T0 | | | | |
|---|---|---|---|---|---|
| | Total | CG[a] | IG[b] | *t* test value (degrees of freedom) | *P* value |
| Depressive symptoms, mean (SD) | 7.26 (5.72) | 7.71 (5.59) | 6.81 (5.82) | 2.031 (671.8) | .04 |
| Anxiety symptoms, mean (SD) | 6.10 (4.96) | 6.52 (4.84) | 5.68 (5.06) | 2.169 (653.9) | .03 |
| Insomnia symptoms, mean (SD) | 3.52 (2.60) | 3.63 (2.61) | 3.41 (2.59) | 1.106 (645.3) | .27 |
| Work stress, mean (SD) | 10.42 (9.75) | 10.37 (9.54) | 10.46 (9.94) | −0.112 (616.9) | .91 |
| General stress, mean (SD) | 6.00 (3.24) | 6.39 (3.11) | 5.61 (3.32) | 3.104 (648.6) | .002 |
| Well-being, mean (SD) | 49.44 (21.47) | 46.54 (21.41) | 52.27 (21.29) | −3.412 (640.1) | .001 |
| MH Quality of Life, mean (SD) | 13.88 (3.32) | 13.59 (3.37) | 14.15 (3.25) | −2.166 (634.6) | .03 |
| Momentaneous well-being, mean (SD) | 6.56 (1.97) | 6.36 (2.03) | 6.74 (1.90) | −2.438 (631.4) | .015 |
| Physical activity, mean (SD) | 6903.26 (6313.60) | 6461.78 (5998.93) | 7388.73 (6619.26) | −1.776 (566.9) | .08 |
| Sitting, mean (SD) | 405.13 (248.54) | 407.85 (241.78) | 402.30 (255.70) | 0.294 (689.6) | .77 |
| Somatic complaints, mean (SD) | 7.66 (4.32) | 8.00 (4.32) | 7.32 (4.30) | 1.989 (632.4) | .047 |
| Absenteeism, mean (SD) | 717.88 (1803.00) | 641.90 (1684.23) | 791.01 (1910.14) | −1.040 (623.6) | .30 |
| Presenteeism, mean (SD) | 8.39 (19.41) | 10.00 (22.83) | 6.84 (15.28) | 2.037 (535.1) | .04 |
| Unpaid labor, mean (SD) | 8.98 (33.40) | 10.58 (38.66) | 7.44 (27.37) | 1.174 (553) | .24 |

[a]CG: control group.

[b]IG: intervention group.

## Adjusted Main Model for Postintervention Effects

All analyses were conducted following an intention-to-treat approach, including all participants according to their original group allocation. The results from models for the different outcomes adjusted by age, gender, and country are displayed in Table 4. The displayed value is the regression coefficient for the interaction effect from group (control vs intervention) by time (pre- to postintervention), therefore representing the intervention effect. Overall, the intervention had no significant effect on any of the outcomes of this study (all $P>.05$). Moreover, due to the significant differences at baseline between the intervention and control group, we also conducted the analyses adjusting for baseline levels. These additional analyses led to similar results except for a lower well-being in the intervention group as compared to the control group ($\beta=-3.289$, $P=.034$, 95% CI −6.331 to −0.247); however, this reduction in well-being is not considered clinically relevant.



XSL•FO
RenderX

**Table .** Generalized linear model adjusted results for postintervention effects.

| Outcome | β-value (95% CI)[a] | *P* value |
|---|---|---|
| Depressive symptoms | −0.052 (−1.010 to 0.905) | .92 |
| Anxiety symptoms | −0.328 (−1.168 to 0.512) | .44 |
| Insomnia symptoms | −0.196 (−0.633 to 0.240) | .38 |
| General stress | 0.385 (−0.195 to 0.965) | .19 |
| Work stress | 0.526 (−1.656 to 2.708) | .64 |
| Well-being | −3.274 (−6.974 to 0.425) | .42 |
| MH[b] Quality of Life | −0.221 (−1.016 to 0.574) | .59 |
| Momentaneous well-being | 0.013 (−0.443 to 0.470) | .95 |
| Physical activity | −224.285 (−2346.582 to 1898.012) | .84 |
| Sitting | −50.343 (−124.856 to 25.169) | .19 |
| Somatic complaints | 0.289 (−0.389 to 0.947) | .39 |
| Absenteeism | −55.672 (−439.783 to 328.439) | .78 |
| Presenteeism | 2.186 (−2.424 to 6.796) | .35 |
| Unpaid labor | −0.04 (−8.356 to 3.653) | .99 |

[a]Adjusted by sex, age, and country.

[b]MH: mental health.

## Adjusted Main Model for Follow-Up Effects

As there was no limit to when the participants could respond to the follow-up questionnaire, the follow-up time had a mean value of 48.84 (SD 24.77) weeks. When looking at the intervention effect at follow-up, no significant intervention effects were found for any of the outcome variables (all *P*>.05; see Table 5). Again, due to the significant differences at baseline between the intervention and control group, we conducted the analyses adjusting for baseline (values or levels), finding no significant differences between the intervention and control group at follow-up.

**Table .** Generalized linear model adjusted results for follow-up effects.

| Outcome | β-value (95% CI)[a] | *P* value |
|---|---|---|
| Depressive symptoms | 0.202 (−0.840 to 1.245) | .70 |
| Anxiety symptoms | 0.375 (−0.537 to 1.287) | .42 |
| Insomnia symptoms | −0.148 (−0.619 to 0.323) | .54 |
| General stress | 0.123 (−0.502 to 0.749) | .70 |
| Work stress | −0.610 (−2.387 to 1.166) | .50 |
| Well-being | −1.022 (−5.019 to 2.975) | .62 |
| MH[b] Quality of Life | −0.211 (−0.945 to 0.524) | .57 |
| Momentaneous well-being | −0.162 (−0.583 to 0.259) | .45 |
| Physical activity | 404.012 (−1219.670 to 2127.693) | .65 |
| Sitting | −34.104 (−96.639 to 28.432) | .28 |
| Somatic health | 0.535 (−0.179 to 1.249) | .14 |
| Absenteeism | −76.124 (−489.616 to 337.368) | .72 |
| Presenteeism | 1.294 (−3.608 to 6.396) | .58 |
| Unpaid labor | 3.653 (−5.322 to 12.629) | .42 |

[a]Adjusted by sex, age, and country.

[b]MH: mental health.



XSL•FO
RenderX

### Results From Subgroup Analyses

Analyses conducted separately by subgroups yielded several significant results (tables with all the outcome variables for the different subgroup analyses can be found in Tables S3-S8 in Multimedia Appendix 1).

First, comparing men and women, we only found that the intervention group had decreased well-being at postintervention (β=−4.796, *P*=.029, 95% CI −9.109 to −0.482) in women. No significant intervention effects were found in men. The subgroup analysis by age groups showed that participants aged 18 - 35 years were less sedentary, spending less time sitting at postintervention (β=283.325, *P*=.004, 95% CI −477.588 to −89.062). Participants aged 50 years and older showed increased presenteeism at postintervention (β=10.016, *P*=.007, 95% CI 2.714 to 17.317). Comparing participants with low and high depression scores at baseline showed that participants with low depressive levels had decreased well-being (β=−4.901, *P*=.029, 95% CI −9.304 to −0.497) and lower mental health quality of life (β=−0.964, *P*=.028, 95% CI −1.822 to −0.105) at postintervention. Conversely, participants with high depressive symptoms at baseline spent less time sitting at T1 (β=−147.023, *P*=.047, 95% CI −292.164 to −1.882). As for low and high levels of anxiety, we observed that those with low anxiety levels at baseline also experienced decreased well-being (β=−4.523, *P*=.048, 95% CI −9.011,−0.036) and reduced mental health quality of life (β=−1.246, *P*=.007, 95% CI −2.148 to −0.344) at T1. Participants with high anxiety levels had increased somatization at follow-up (β=1.413, *P*=.025, 95% CI 0.181 to 2.646). As for low and high levels of somatization at baseline, participants with low somatization at baseline were found to have a decrease in mental health quality of life at T1 (β=−2.095, *P*=.003, 95% CI −3.498 to −0.692), whereas they experienced less insomnia symptoms at follow-up (β=−0.855, *P*=.026, 95% CI −1.606 to −0.103). Participants with high levels of somatization at baseline had decreased well-being (β=−4.489, *P*=.049, 95% CI −8.962 to −0.017) and spent less time sitting at T1 (β=−103.532, *P*=.029, 95% CI −196.746 to −10.318). Finally, considering low and high levels of work stress, individuals with low work stress were found to report less well-being at follow-up (β=−6.673, *P*=.042, 95% CI −13.102 to −0.244), while individuals with high work stress reported less insomnia at follow-up (β=−0.822, *P*=.024, 95% CI −1.534 to −0.110).

### Attrition Analysis

The attrition analysis revealed several significant baseline differences between the participants who dropped out from this study and those who did not drop out (Table S9 in Multimedia Appendix 1). Age was slightly lower in the dropout group (mean 44.17, SD 10.03) compared to the not-dropout group (mean 44.23, SD 11.01), with a significant difference (*t*=3.201, *P*=.001). There were also significant differences in country distribution ($\chi^2$=13.135, *P*=.004), with a higher proportion of participants from the United Kingdom and a lower proportion from Spain and Poland in the dropout group. Depressive symptoms, anxiety symptoms, general stress, and work stress levels were significantly higher in the dropout group (depressive symptoms: *t*=−2.092, *P*=.037; anxiety symptoms: *t*=−2.784, *P*=.005; general stress: *t*=−3.101, *P*=.002; work stress: *t*=−2.133, *P*=.033). Absenteeism was also higher in the dropout group (*t*=−2.942, *P*=.003). Other variables did not show significant differences between the groups. Moreover, of the participants in the intervention group, only 34% were considered active users of the intervention, implying an overall low intervention adherence.

### Per-Protocol Analyses

Due to high dropout rates and low engagement, we performed additional analyses to assess the impact of those biases. We divided participants in the intervention group into active and inactive users depending on their level of interaction with the application. We conducted LMM analyses and found that, again, none of the outcome measures was significantly different from the control group for both inactive and active groups, and for postintervention and follow-up measurement time (all *P*>.05, Table S10 in Multimedia Appendix 1). Additionally, when comparing active participants to inactive participants, no significant differences in any of the outcomes were found (all *P*>.05, Table S11 in Multimedia Appendix 1).

### Sample Size and Sensitivity Analysis

The final sample comprised 708 participants (347 in the intervention and 361 in the control group). Considering the corrective factor of 1.2 used in the a priori calculation (to account for the cluster design), the effective sample sizes are approximately 289 and 303 per group. With these numbers, the minimum detectable effect size for 80% power and 2-sided α=0.05 is 0.24-0.25 [44]. Thus, the current study would show adequate power to detect small-to-moderate effects (≥0.25-0.30) but would be underpowered for very small effects (≤0.20).

In a per-protocol comparison between intervention users (n=119) and controls (n=363), the effective intervention sample would drop to approximately 99 participants after the corrective factor, yielding a minimum detectable effect size of *d* between 0.34-0.35. This would considerably reduce the statistical power to detect small effects and should therefore be taken into account when interpreting our null findings.

### Harms or Unintended Effects

No adverse events or unintended effects related to the intervention were reported.

## Discussion

### Principal Findings

In this cluster RCT, we investigated whether a multimodal, digital platform, consisting of a website and an application, improved mental health, well-being, and work-related outcomes of employees from SMEs and public agencies in 4 European countries.

### Effectiveness Results

Overall, we did not find evidence of effectiveness of the intervention in the short (7 wk) or long term (21 wk post intervention), but some positive effects were observed for certain outcomes and subgroups of employees. We did not observe the expected effectiveness of the intervention to improve depressive



XSL•FO
RenderX

symptoms, anxiety symptoms, general and work-related stress, insomnia symptoms, physical activity, well-being, presenteeism, or absenteeism. However, different effect patterns were found when analyzing different subgroups. Previous literature showed small, significant effects of digital interventions in the workplace [9]. In the Carolan et al [9] meta-analysis (2017), 9 of the 21 studies included did not find an effect of the digital intervention on mental health. Additionally, 8 of 13 included studies did not find an intervention effect on work outcomes. It is possible that longer follow-up assessments (more than 21 weeks) are needed to detect an effect of an intervention like that proposed by EMPOWER.

As previously stated, when analyzing the effectiveness of the intervention in different subgroups, we found some significant results. For younger participants (18 - 35 y), those with high levels of depressive symptoms at baseline and with high baseline somatization, the intervention showed a positive effect to reduce sedentary behavior. Additionally, insomnia improved at follow-up for participants with low baseline somatization. However, in other subgroups, the intervention had the opposite effect. These findings suggest that demographic factors and initial health conditions could significantly influence the effectiveness and outcomes of interventions, highlighting the need for tailored approaches in mental health and well-being programs. Previous evidence has shown that tailored digital interventions did not reduce anxiety and depressive symptoms in the general working population, but significantly alleviated these conditions in employees with higher psychological distress [12]. More research is needed to determine which characteristics of employees could moderate the effectiveness of OeMH interventions.

Contrary to our expectations, some subgroups (eg, women, participants with low baseline depressive and anxiety symptoms, and those with high baseline somatization) showed a short-term decrease in well-being post intervention, which disappeared at follow-up. Temporary declines in well-being and mental health after a psychological intervention are not uncommon and could result from increased self-awareness of one's own issues and challenges. This is in line with the prevalence inflation hypothesis, which suggests that greater introspection and symptom monitoring can increase perceived distress before improvement occurs [45].

Indeed, the prevalence inflation hypothesis posits that heightened awareness during interventions can lead to increased reporting of mild symptoms as significant issues, potentially exaggerating perceived prevalence. According to this, while mental health awareness efforts are beneficial as they lead to more accurate reporting of previously unrecognized symptoms, these efforts may also cause individuals to interpret and report mild distress as mental health problems, potentially exacerbating symptoms through self-fulfilling behaviors and creating a cyclical and intensifying effect [45]. One could speculate that immediate benefits of an intervention such as EMPOWER might not always be observable and thus, longer follow-up assessments are needed to observe positive effects.

A recent review [12] suggests that personalized digital interventions demonstrated positive outcomes when it comes to reducing presenteeism, alleviating stress, improving sleep, and addressing physical symptoms associated with somatization. Nevertheless, their effectiveness in reducing absenteeism is somewhat limited. Moreover, another systematic review reported that the content of the intervention based on stress management and mindfulness seems more effective than CBT-based content. This suggests that CBT may be less useful for preventive approaches and for workplace settings [46]. All in all, the findings suggest that a universal approach to digital mental health interventions may not be effective in addressing the nuanced needs of workplace populations. Future implementations should consider a targeted approach, focusing on individuals with expressed mental health challenges or higher baseline distress.

## Challenges During the RCT

Given the various challenges we faced during the trial, we cannot dismiss the possibility that the EMPOWER results may be attributable to chance rather than reflecting the effectiveness of the EMPOWER intervention. Some of these challenges included external contextual factors, such as the COVID-19 pandemic or the Ukraine war, organizational barriers, complexity of the initial design and length of questionnaires, and technical problems with the application affecting accessibility and adherence.

On the one hand, several contextual challenges were encountered during the RCT. For instance, in Finland, there were large reforms of health care, social welfare, and rescue services at the beginning of 2023. In Poland, there was a lack of company involvement in promoting the project among employees and a fear of discussing mental health with employees. In the United Kingdom, there were several strikes, and in Spain, we observed a certain level of stigma in the workplace to openly talk about mental health in the workplace. Overall, the COVID-19 pandemic hindered the successful completion of this trial, and the Ukraine war has had a significant impact on the small and medium companies both in Finland and in Poland.

On the other hand, we faced several challenges related to this study's design and the technical problems of the application itself. During the design phase, we conducted a prepilot study with volunteers who tested the beta version of the application [23]. However, there was no proper validation of the prototype in a real-world context (ie, employees), which hindered the early identification of technical problems with the design and functionality of the application. As such, during the RCT, users reported software bugs of the application, which interfered with the participants' accessibility to the intervention. However, we did not collect information on whether users who reported malfunctioning of the application were less engaged with the intervention. Several users contacted local researchers to report a bug in the registration process. To solve it, we provided them with a 1-use link to register. This bug could not be fixed, and we believe this could have also led to a loss of potential participants who did not contact us.

Moreover, the first study design (stepped-wedge design) required participants to answer many questionnaires at specific times. This resulted in a burdensome task, especially for those workers on the waiting list who complained about the number



XSL•FO
RenderX

of questionnaires with no obvious positive return. The design was then modified to a 2-parallel cluster control trial in September 2022, which helped increase the number of participants and retention. Additionally, invitation emails were initially sent by employers through a specific back-end system. We observed that this method, despite providing us with more control over who was participating in this study, did not result in an optimal recruitment rate. Thus, we created flyers with QR codes to directly download the application. While these changes were necessary to address low recruitment and high attrition, they introduced additional complexity in the interpretation of findings.

Only 34.3% of participants from the intervention group actively interacted with the application. The application contained a simple gamification system (ie, medals) that might not be enough to achieve an optimal engagement. During the qualitative data collection, participants indicated that the application was easy to use and enjoyable, but there was too much text and suggested the convenience of using more interactive and visual content and more feedback about one's progress. Additionally, the application was designed as a web application, which could introduce some problems in installing and receiving push-up notifications. Dropout rates were also high in our study, especially at T1 (postintervention) with 49% and 57% of loss to follow-up in the intervention and control group, respectively. Interestingly, the rate of retention was higher for the last round of questionnaires, with 70% from the intervention group and 91.7% from the control group ending questionnaires at follow-up. This could have affected the statistical power needed to find significant effects of the intervention.

### Lessons Learned and Future Perspectives

During this study's process, we collected information about several challenges that provided a better understanding of barriers and facilitators of OeMH in the workplace. As such, we believe the lessons learned from the EMPOWER project can provide critical insights for future adaptations and evaluations of similar interventions.

Concerning this study's design, 1 significant challenge was ensuring that both employees and employers understood that EMPOWER was still in the research phase and not a finalized implementation. This issue was compounded by the need to balance strict RCT protocols with practical workplace application, a tension exacerbated by disruptions from the COVID-19 pandemic and the war in Ukraine's impact on neighboring countries such as Finland and Poland. Emergency protocols would have facilitated the recruitment process and adherence to the intervention. Flexibility in research design was crucial to completing the RCT amid unforeseen circumstances, preserving the scientific integrity of the effectiveness evaluation, and underscoring the importance of contingency planning in complex settings. Indeed, a restrictive RCT with multiple repeated measures may not be the most effective method for evaluating educational and skills-based interventions aimed at improving mental health and quality of life in the general worker population. Related to this, flexible trial approaches, including adaptive trial designs, have been suggested as alternatives to the traditional fixed-design trial. They allow for prospectively planned modifications to one or more aspects of the design based on data accumulated over the course of this trial as part of this study's protocol [47] and are thought to offer several benefits compared to the traditional fixed designs [48]. These benefits include greater statistical efficiency, which can be considered more ethical in some cases, and can be used to answer broader questions and increase acceptance from participants and trial funders. On the other hand, flexible trials can also add higher complexity because they require more planning and methodological considerations [48]. They should preserve statistical rigor, ensure transparency, and add prespecifications at the planning phase of the current study.

Another barrier was the omission of direct involvement from actual end users (managers and employees) in the design phase, highlighting the need for a more inclusive, bottom-up co-design process. However, early engagement with a broad range of stakeholders, including labor physicians, psychologists, union representatives, and HR managers, was initially beneficial [30]. Future implementations should follow a multistep approach to readiness that could integrate the development of the proof of concept, prototyping, and validation, piloting in a relevant environment, demonstration in real-world settings, and release of the tool, as suggested in the technology readiness level framework [32]. It is important to collect direct input from real-world end users, tailoring the tool to their specific contexts and needs using a co-design approach. In this context, it is crucial to assess the organizational readiness before OeMH interventions, including technological capacity and capability, commitment from both employers and employees, management support, and attitudes toward mental health. Implementing a brief, systematic, and standardized workplace assessment, such as the EMPOWER checklist [49], can help identify and address key threats to the RCT or pilot implementations. This could facilitate recruitment and adherence to the intervention, setting the stage for future uptake.

Privacy concerns were another identified barrier in our study. Indeed, concerns regarding privacy and data security necessitate clear and transparent privacy policies detailing data collection, usage, storage, and employees' rights, ensuring employers have no access to personal data. Developing a brief and structured guide for researchers and participants without a legal background can help identify privacy and security red flags early on.

The low adherence observed in our study may also reflect a poor user experience with the application and technical problems. Adherence to self-guided apps for depression and anxiety has been reported to range from 7% to 42% after 4 weeks, and 0.5% - 28.6% after 6 weeks [50]. In our study, some barriers mentioned by employees included a lack of time to use it, long questionnaires, unclear instructions of how to use it, and long text, whereas the application was referred to as easy to use and enjoyable. Several elements have been suggested to increase adherence to OeMH interventions, such as offering tailored and interactive content, peer support, monitoring and feedback about one's emotional state, and the possibility of contact with a coach [7].



XSL•FO
RenderX

Finally, policy support should focus on the continuous improvement and testing of existing platforms, such as the EMPOWER eHealth platform, rather than developing new tools. The lessons learned and infrastructure developed during the project provide a strong basis for further development, which is more sustainable and cost-effective compared to creating new tools from scratch. By following these recommendations, future OeMH interventions can be more effectively implemented and evaluated, ensuring better outcomes for both organizations and employees.

### Study Limitations

As previously mentioned, our results should be interpreted cautiously due to the several challenges encountered during the RCT. Additional limitations that need to be considered are, first, that this study aimed for a conservative target sample size of 874 participants. Although 877 employees agreed to participate and 81% (n=708) answered the baseline assessments, only 333 (47%) and 249 (35.2%) answered the follow-up assessments. Attrition bias is generally a concern in the case of OeMH interventions, with unsupervised interventions having a mean of 32.85% (SD 22.21%) of attrition at postintervention [51]. Thus, selection and attrition bias could have played a role too in the effectiveness results, as there were significant differences in the baseline characteristics of participants who dropped out of this study compared to those who did not. Second, adherence to the EMPOWER intervention was notably low, with only about 34% considered as active users. Comparison of active and inactive users with the control group did not show any positive intervention effects. Given that only 1 in 3 participants continued using the intervention after 7 weeks, it is probable that the observed results are due to chance rather than the effectiveness of the EMPOWER intervention and do not reflect the potential of the intervention to improve mental health problems in the workplace. Third, because randomization was performed at the cluster level (companies or departments) rather than the individual level, this study was vulnerable to baseline imbalances between groups. Indeed, the intervention group showed significantly lower levels of depression, anxiety, and stress, and higher levels of well-being and quality of life at baseline. These differences may have limited the scope for observable improvement and introduced bias despite adjustment in analyses. Moreover, the relatively modest number of clusters (25 in the intervention arm and 19 in the control arm) increased the likelihood of imbalance. As such, the randomization process might not have succeeded in creating 2 comparable samples, adding to the potential floor effect. This can often happen in clustered RCT's, as when not having a large number of individuals to randomize but rather a limited number of clusters, the result of the randomization is more prone to cause imbalances between trial arms [24]. In our study, baseline imbalances between intervention and control groups were observed, with the intervention group showing significantly lower levels of depression, anxiety, and stress, and higher levels of well-being and mental health quality of life. These differences could potentially introduce bias, as the intervention group's baseline characteristics may limit the scope for observable improvements. Additionally, our failure to find significant differences between groups could be due to the regression to the mean effect [52]. To address this, all models were adjusted for baseline levels of these variables, and subgroup analyses were conducted to further account for these imbalances. These subgroup analyses revealed some significant results, such as the fact that women in the intervention group had worse well-being at postintervention, or that participants with higher stress levels at baseline had better insomnia scores in the follow-up after using the intervention. Additionally, baseline-adjusted analyses reflected similar results to the ones from unadjusted models. Despite these efforts, the imbalances underscore the challenges of achieving comparable groups in cluster randomized trials, particularly when working with smaller samples and diverse organizational contexts. Fourth, the modification from a stepped-wedge design to a 2-arm parallel cluster RCT, while approved by ethics committees and justified by recruitment and attrition challenges, may have further complicated comparability between groups. Participants initially enrolled under the stepped-wedge design were reassigned to the new trial structure, which could have affected trial integrity and interpretation of results. Fifth, the use of self-reported measures could introduce response bias and thus also be more prone to the prevalence inflation hypothesis. Sixth, there were statistical differences at baseline between both the intervention and control groups in demographic and outcome measures, which could have added bias to the results. Seventh, it is also important to interpret subgroup analyses with caution, as wide CIs for effect estimates indicate variability and uncertainty. These findings emphasize the need for tailored strategies and robust feasibility assessments in future trials. Eighth, as the intervention was delivered in a nonclinical population, individuals in a generally healthy population tend to start with lower initial levels of psychological symptoms, which means there is limited potential for further reduction [53]. As a result, the minimal observed improvements in these symptoms may be explained by a floor effect, where the initial low scores leave little room for noticeable change ("floor effect" refers to a scenario where baseline scores on outcome measures are so low that further reductions are difficult to detect, limiting the potential for improvement). Related to this, measuring significant changes in clinical outcomes (such as depression and anxiety symptoms) within a predominantly healthy population can be challenging. Preventive interventions often yield small effect sizes, and detecting statistically significant changes requires large sample sizes and sensitive measurement tools [54]. Finally, although the initial sample size calculation was conservative, the final effective sample size was lower than planned. The study had sufficient power to detect small-to-moderate effects ($d \geq 0.25$-$0.30$), but not very small effects ($d \leq 0.20$). Moreover, adherence was low, with only 119 of 347 participants in the intervention group actively using the program. This reduced the power of per-protocol analyses and likely diluted the effect sizes observed under ITT. Therefore, null results should be interpreted with caution, as this study cannot rule out the presence of smaller effects.

### Conclusion

This study represented a large, multicountry cluster RCT evaluating the effectiveness of the EMPOWER platform, a digital OeMH intervention. Its main contribution lies in the



XSL•FO
RenderX

rigorous assessment of both effectiveness and implementation challenges across diverse organizational contexts, particularly within SMEs and public agencies. In contrast to many studies conducted in more controlled or clinical samples, EMPOWER targeted a predominantly healthy working population, highlighting the difficulties of detecting effects in preventive workplace interventions.

While the intervention did not demonstrate significant effects on primary or secondary outcomes, the findings provide valuable guidance for future research and practice by emphasizing the importance of feasibility testing, organizational readiness, and strategies to enhance user engagement. From a real-world perspective, the results support the need for conducting comprehensive feasibility studies, ensuring organizational readiness, and adopting flexible trial designs to enhance engagement and adherence in diverse workplace settings.

## Acknowledgments

During the preparation of this work, a generative artificial intelligence (AI) tool (ChatGPT, OpenAI) was used solely to check the grammatical accuracy of selected parts of the manuscript. The authors reviewed all outputs and take full responsibility for the final content.

## Funding

This project has received funding from the European Union's Horizon 2020 research (https://cordis.europa.eu/project/id/848180) and innovation program under grant agreement No 848180, and the National Health and Medical Research Council (NHMRC) of Australia under Grant Agreement APP1195937. CdM has received funding in form of a predoctoral grant from the Generalitat de Catalunya (PIF-Salut grant, code SLT017/20/000138). AR is supported by the Italian Ministry of Health (RRC). BO is supported by the Miguel Servet (CP20/00040) contract, funded by the Instituto de Salud Carlos III (Spain). The funders had no involvement in this study's design, data collection, analysis, interpretation, or the writing of this paper.

## Data Availability

Requests to access the dataset should be directed to the corresponding author.

## Authors' Contributions

BO and CdM handled the analysis, interpretation, and paper writing. JR-B worked on the analysis. All authors have engaged in paper drafting and have approved the final version for publication.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Tables of data material.
[DOCX File, 115 KB - i-jmr_v15i1e66041_app1.docx ]

Checklist 1
CONSORT checklist.
[PDF File, 331 KB - i-jmr_v15i1e66041_app2.pdf ]

## References

1. EU-compass for action on mental health and well-being. European Commission. URL: https://health.ec.europa.eu/non-communicable-diseases/mental-health/eu-compass-action-mental-health-and-well-being_en [accessed 2026-03-06]
2. Leka S, Jain A. Mental health in the workplace in europe. EU Compass for Action on Mental Health and Well-Being. URL: https://ec.europa.eu/health/system/files/2017-07/compass_2017workplace_en_0.pdf [accessed 2026-03-06]
3. WHO fact sheet: mental health at work. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/mental-health-at-work [accessed 2026-03-06]
4. Olesen J, Gustavsson A, Svensson M, et al. The economic cost of brain disorders in Europe. Eur J Neurol 2012 Jan;19(1):155-162. [doi: 10.1111/j.1468-1331.2011.03590.x] [Medline: 22175760]
5. Brouwer WBF, Koopmanschap MA, Rutten FFH. Productivity costs in cost‐effectiveness analysis: numerator or denominator: a further discussion. Health Econ 1997 Sep;6(5):511-514. [doi: 10.1002/(SICI)1099-1050(199709)6:5<511::AID-HEC297>3.3.CO;2-B]
6. Kigozi J, Jowett S, Lewis M, Barton P, Coast J. The estimation and inclusion of presenteeism costs in applied economic evaluation: a systematic review. Value Health 2017 Mar;20(3):496-506. [doi: 10.1016/j.jval.2016.12.006]



XSL•FO
RenderX

7. Carolan S, de Visser RO. Employees' perspectives on the facilitators and barriers to engaging with digital mental health interventions in the workplace: qualitative study. JMIR Ment Health 2018 Jan 19;5(1):e8. [doi: 10.2196/mental.9146]
8. Miguel C, Amarnath A, Akhtar A, et al. Universal, selective and indicated interventions for supporting mental health at the workplace: an umbrella review of meta-analyses. Occup Environ Med 2023 Apr;80(4):225-236. [doi: 10.1136/oemed-2022-108698] [Medline: 36828633]
9. Carolan S, Harris PR, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of web-based psychological interventions delivered in the workplace. J Med Internet Res 2017 Jul 26;19(7):e271. [doi: 10.2196/jmir.7583] [Medline: 28747293]
10. Stratton E, Jones N, Peters SE, Torous J, Glozier N. Digital mHealth interventions for employees: systematic review and meta-analysis of their effects on workplace outcomes. J Occup Environ Med 2021 Aug 1;63(8):e512-e525. [doi: 10.1097/JOM.0000000000002267] [Medline: 34050095]
11. Stratton E, Lampit A, Choi I, Calvo RA, Harvey SB, Glozier N. Effectiveness of eHealth interventions for reducing mental health conditions in employees: a systematic review and meta-analysis. PLoS One 2017;12(12):e0189904. [doi: 10.1371/journal.pone.0189904] [Medline: 29267334]
12. Moe-Byrne T, Shepherd J, Merecz-Kot D, et al. Effectiveness of tailored digital health interventions for mental health at the workplace: a systematic review of randomised controlled trials. PLOS Digit Health 2022 Oct;1(10):e0000123. [doi: 10.1371/journal.pdig.0000123] [Medline: 36812547]
13. Corbière M, Shen J, Rouleau M, Dewa CS. A systematic review of preventive interventions regarding mental health issues in organizations. Work 2009;33(1):81-116. [doi: 10.3233/WOR-2009-0846] [Medline: 19597288]
14. Bhui KS, Dinos S, Stansfeld SA, White PD. A synthesis of the evidence for managing stress at work: a review of the reviews reporting on anxiety, depression, and absenteeism. J Environ Public Health 2012;2012:515874. [doi: 10.1155/2012/515874] [Medline: 22496705]
15. Purtova N, Kosta E, Koops BJ. Laws and regulations for digital health. In: Requirements Engineering for Digital Health: Springer; 2015:47-74. [doi: 10.1007/978-3-319-09798-5_3]
16. Bernard RM, Toppo C, Raggi A, et al. Strategies for implementing occupational eMental health interventions: scoping review. J Med Internet Res 2022 Jun 1;24(6):e34479. [doi: 10.2196/34479] [Medline: 35648457]
17. Sevic A, Hashemi NS, Thørrisen MM, et al. Effectiveness of eHealth interventions targeting employee health behaviors: systematic review. J Med Internet Res 2023 Apr 20;25:e38307. [doi: 10.2196/38307] [Medline: 37079369]
18. Clement S, Schauman O, Graham T, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychol Med 2015 Jan;45(1):11-27. [doi: 10.1017/S0033291714000129] [Medline: 24569086]
19. Wang-Schweig M, Kviz FJ, Altfeld SJ, Miller AM, Miller BA. Building a conceptual framework to culturally adapt health promotion and prevention programs at the deep structural level. Health Promot Pract 2014 Jul;15(4):575-584. [doi: 10.1177/1524839913518176] [Medline: 24396122]
20. Benning FE, van Oostrom SH, van Nassau F, Schaap R, Anema JR, Proper KI. The implementation of preventive health measures in small- and medium-sized enterprises—a combined quantitative/qualitative study of its determinants from the perspective of enterprise representatives. IJERPH 2022;19(7):3904. [doi: 10.3390/ijerph19073904]
21. Aust B, Leduc C, Cresswell-Smith J, et al. The effects of different types of organisational workplace mental health interventions on mental health and wellbeing in healthcare workers: a systematic review. Int Arch Occup Environ Health 2024 Jul;97(5):485-522. [doi: 10.1007/s00420-024-02065-z] [Medline: 38695906]
22. Olaya B, Van der Feltz-Cornelis CM, Hakkaart-van Roijen L, et al. Study protocol of EMPOWER: a cluster randomized trial of a multimodal eHealth intervention for promoting mental health in the workplace following a stepped wedge trial design. Digit Health 2022;8:20552076221131145. [doi: 10.1177/20552076221131145] [Medline: 36276189]
23. van der Feltz-Cornelis CM, Shepherd J, Gevaert J, et al. Design and development of a digital intervention for workplace stress and mental health (EMPOWER). Internet Interv 2023 Dec;34:100689. [doi: 10.1016/j.invent.2023.100689] [Medline: 38054076]
24. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340(mar23 1):c332-c332. [doi: 10.1136/bmj.c332]
25. Campbell MK, Piaggio G, Elbourne DR, Altman DG, Group C. CONSORT 2010 statement: extension to cluster randomised trials. BMJ 2012 Sep 4;345:e5661. [doi: 10.1136/bmj.e5661] [Medline: 22951546]
26. Hopewell S, Clarke M, Moher D, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. PLoS Med 2008 Jan 22;5(1):e20. [doi: 10.1371/journal.pmed.0050020] [Medline: 18215107]
27. Esserman D, Allore HG, Travison TG. The method of randomization for cluster-randomized trials: challenges of including patients with multiple chronic conditions. Int J Stat Med Res 2016 Jan 8;5(1):2-7. [doi: 10.6000/1929-6029.2016.05.01.1] [Medline: 27478520]
28. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016. : European Union; 2016 Apr 27 URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng [accessed 2026-03-06]




XSL•FO
RenderX

29. Moerbeek M, Van Breukelen GJP, Berger MPF. Optimal experimental designs for multilevel logistic models. J R Stat Soc D 2001 Mar;50(1):17-30. [doi: 10.1111/1467-9884.00257]
30. Chisholm D, Sweeny K, Sheehan P, et al. Scaling-up treatment of depression and anxiety: a global return on investment analysis. Lancet Psychiatry 2016 May;3(5):415-424. [doi: 10.1016/S2215-0366(16)30024-4] [Medline: 27083119]
31. Salvador-Carulla L, Woods C, de Miquel C, Lukersmith S. Adaptation of the technology readiness levels for impact assessment in implementation sciences: the TRL-IS checklist. Heliyon 2024 May 15;10(9):e29930. [doi: 10.1016/j.heliyon.2024.e29930] [Medline: 38726201]
32. Salvador-Carulla L, Lukersmith S, Woods CE, Chen T, de Miquel C. A protocol using mixed methods for the impact analysis of the implementation of the EMPOWER project: an eHealth intervention to promote mental health and well-being in European workplaces. BMJ Open 2025 Apr 9;15(4):e082219. [doi: 10.1136/bmjopen-2023-082219] [Medline: 40204313]
33. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
34. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
35. Bouwmans C, Krol M, Severens H, Koopmanschap M, Brouwer W, Hakkaart-van Roijen L. The iMTA productivity cost questionnaire: a standardized instrument for measuring and valuing health-related productivity losses. Value Health 2015 Sep;18(6):753-758. [doi: 10.1016/j.jval.2015.05.009] [Medline: 26409601]
36. Bastien CH, Vallières A, Morin CM. Validation of the insomnia severity index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]
37. Najder A, Merecz-Kot D, Wójcik A. Relationships between occupational functioning and stress among radio journalists--Assessment by means of the psychosocial risk scale. Int J Occup Med Environ Health 2016;29(1):85-100. [doi: 10.13075/ijomeh.1896.00332] [Medline: 26489946]
38. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404] [Medline: 6668417]
39. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176. [doi: 10.1159/000376585] [Medline: 25831962]
40. van Krugten FCW, Busschbach JJV, Versteegh MM, Hakkaart-van Roijen L, Brouwer WBF. The Mental Health Quality of Life Questionnaire (MHQoL): development and first psychometric evaluation of a new measure to assess quality of life in people with mental health problems. Qual Life Res 2022 Feb;31(2):633-643. [doi: 10.1007/s11136-021-02935-w] [Medline: 34241821]
41. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
42. Kroenke K, Spitzer RL, Williams JBW. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 2002;64(2):258-266. [doi: 10.1097/00006842-200203000-00008] [Medline: 11914441]
43. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: intention-to-treat versus per-protocol analysis. Perspect Clin Res 2016;7(3):144-146. [doi: 10.4103/2229-3485.184823] [Medline: 27453832]
44. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Lawrence Erlbaum Associates; 1988. URL: https://utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf [accessed 2026-03-28]
45. Foulkes L, Andrews JL. Are mental health awareness efforts contributing to the rise in reported mental health problems? A call to test the prevalence inflation hypothesis. New Ideas Psychol 2023 Apr;69:101010. [doi: 10.1016/j.newideapsych.2023.101010]
46. Stratton E, Lampit A, Choi I, et al. Trends in effectiveness of organizational eHealth interventions in addressing employee mental health: systematic review and meta-analysis. J Med Internet Res 2022 Sep 27;24(9):e37776. [doi: 10.2196/37776] [Medline: 36166285]
47. Pallmann P, Bedding AW, Choodari-Oskooei B, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. BMC Med 2018 Feb 28;16(1):29. [doi: 10.1186/s12916-018-1017-7] [Medline: 29490655]
48. Tricco AC, Tovey D. Flexible approaches to clinical trials. J Clin Epidemiol 2023 Feb;154:A1-A2. [doi: 10.1016/j.jclinepi.2023.02.004] [Medline: 36933905]
49. Raggi A, Bernard RM, Toppo C, et al. The EMPOWER occupational e-mental health intervention implementation checklist to foster e-Mental health interventions in the workplace: development study. J Med Internet Res 2024 Mar 15;26(1):e48504. [doi: 10.2196/48504] [Medline: 38488846]
50. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
51. de Miquel C, Haro JM, van der Feltz-Cornelis CM, et al. Differential attrition and engagement in randomized controlled trials of occupational mental health interventions in person and online: a systematic review and meta-analysis. Scand J Work Environ Health 2024 Dec 1;50(8):588-601. [doi: 10.5271/sjweh.4173] [Medline: 39072699]
52. Morton V, Torgerson DJ. Regression to the mean: treatment effect without the intervention. J Eval Clin Pract 2005 Feb;11(1):59-65. [doi: 10.1111/j.1365-2753.2004.00505.x] [Medline: 15660538]




XSL•FO
RenderX

53. Bower P, Kontopantelis E, Sutton A, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. BMJ 2013 Feb 26;346:f540. [doi: 10.1136/bmj.f540] [Medline: 23444423]
54. Joyce S, Modini M, Christensen H, et al. Workplace interventions for common mental disorders: a systematic meta-review. Psychol Med 2016 Mar;46(4):683-697. [doi: 10.1017/S0033291715002408] [Medline: 26620157]

## Abbreviations

**CBT:** cognitive behavioral therapy
**CONSORT:** Consolidated Standards of Reporting Trials
**EMPOWER:** The European Platform to Promote Wellbeing and Health in the Workplace
**GAD-7:** Generalized Anxiety Disorder Questionnaire
**LMM:** linear mixed model
**OeMH:** occupational e-mental health
**PHQ-9:** Patient Health Questionnaire-9
**RCT:** randomized controlled trial
**SME:** small to medium enterprise

*Edited by M Balcarras; submitted 02.Sep.2024; peer-reviewed by CY Chen, P Baldwin; revised version received 26.Jan.2026; accepted 26.Jan.2026; published 14.Apr.2026.*

*Please cite as:*
*de Miquel C, Van der Feltz-Cornelis CM, Hakkaart-van Roijen L, Merecz-Kot D, Sinokki M, Rodeiro-Boliart J, Sweetman J, Staszewska K, Vorstenbosch E, Porricelli D, Peeters S, Ayuso-Mateos JL, Salvador-Carulla L, Lukersmith S, Borrega O, Sabariego C, Vanroelen C, Raggi A, Porcheddu D, Haro JM, Olaya B*
*Effectiveness and Lessons Learned From an Occupational E-Mental Health Intervention for Enhancing Workplace Mental Health: The EMPOWER Cluster Randomized Controlled Trial*
*Interact J Med Res 2026;15:e66041*
*URL: https://www.i-jmr.org/2026/1/e66041*
*doi:10.2196/66041*

© Carlota de Miquel, Christina M Van der Feltz-Cornelis, Leona Hakkaart-van Roijen, Dorota Merecz-Kot, Marjo Sinokki, Jordi Rodeiro-Boliart, Jennifer Sweetman, Kaja Staszewska, Ellen Vorstenbosch, Daniele Porricelli, Stijn Peeters, José Luis Ayuso-Mateos, Luis Salvador-Carulla, Sue Lukersmith, Oriol Borrega, Carla Sabariego, Christophe Vanroelen, Alberto Raggi, Diletta Porcheddu, Josep Maria Haro, Beatriz Olaya. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 14.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Perspectives of Indian Gastroenterologists and Hepatologists on Nonalcoholic Fatty Liver Disease Diagnosis and Management: Insights From the Nationwide Web-Based Cross-Sectional DRIVE Survey

Anil Arora[1], DM; Yogesh Garje[2†], MD; Shagupta Shaikh[2], MD, DNB; Shruti Dharmadhikari[2], MSc; Chintan Khandhedia[2], MD; Neeraj Markandeywar[2], MD; Amey Mane[2], MD; Suyog C Mehta[2], MD; Ashish Kumar[1], DM

[1]Institute of Liver, Gastroenterology and Pancreaticobiliary Sciences, Sir Ganga Ram Hospital, Rajinder Nagar, New Delhi, India

[2]Medical Affairs & Clinical Research, Sun Pharma Laboratories Limited, Mumbai, Maharashtra, India

[†]deceased

**Corresponding Author:**
Shagupta Shaikh, MD, DNB
Medical Affairs & Clinical Research
Sun Pharma Laboratories Limited
Sun House, Plot No. 201 B/1
Western Express Highway, Goregaon (East)
Mumbai, Maharashtra, 400063
India
Phone: 91 9766871372
Email: Shagupta.Shaikh@sunpharma.com

## Abstract

**Background:** Nonalcoholic fatty liver disease (NAFLD) and its progressive form, nonalcoholic steatohepatitis (NASH), represent an increasing clinical and public health burden in India. Despite their high prevalence, there are limited data on their diagnostic and management approaches among Indian health care providers. Real-world evidence on how Indian gastroenterologists and hepatologists diagnose and manage these conditions remains limited.

**Objective:** This study aimed to understand the current disease perspectives, diagnostic modalities, and management practices for NAFLD and NASH among Indian hepatologists and gastroenterologists.

**Methods:** A nationwide, web-based cross-sectional survey was conducted online between May 2023 and July 2023 among practicing gastroenterologists and hepatologists from health care setups, clinics, and hospitals located across India. The structured, self-administered questionnaire included 34 items covering 3 domains: disease perspectives (n=16), diagnostic modalities (n=4), and management strategies (n=14). Descriptive statistics were used to summarize responses as counts and percentages.

**Results:** A total of 609 physicians completed the online survey (gastroenterologists: n=556, 91.3%; hepatologists: n=53, 8.7%). For 336 (55.2%) physicians, NAFLD accounted for 25% to 50% of the patients consulted per month, and 220 (36.1%) physicians reported that 10% to 20% of patients with NAFLD had NASH. Obesity (n=583, 95.7%) and diabetes (n=579, 95.1%) were cited as leading risk factors for NAFLD. Transient elastography was the diagnostic tool preferred by 558 (91.6%) physicians, followed by NAFLD fibrosis score (n=378, 62.1%) and Fibrosis-4 score (n=356, 58.5%); only 154 (25.3%) physicians used liver biopsy. For treatment, 414 (68%) physicians managed patients using pharmacotherapy and dietary and lifestyle modifications, while 195 (32%) relied on lifestyle modification alone. Antioxidant vitamins (n=543, 89.2%) and saroglitazar (n=522, 85.7%) were the most frequently prescribed therapies. The main barriers to optimal NASH management reported were lack of patient awareness (n=466, 76.5%) and limited availability of effective pharmacological options (n=303, 49.8%).

**Conclusions:** This large, nationwide survey highlights that NAFLD and NASH constitute a major part of gastroenterology and hepatology practice in India. Although transient elastography and pharmacological agents such as saroglitazar and vitamin E are widely used, considerable heterogeneity exists in diagnostic and management approaches. The lack of patient awareness and effective treatment options remain the major hurdles in managing NAFLD and NASH. These findings underscore the need for the wider implementation of existing India-specific consensus recommendations, continued physician education, and future research focusing on tailored interventions in the management of NAFLD and NASH for the Indian population.



XSL•FO
RenderX

*(Interact J Med Res 2026;15:e75138)* doi:10.2196/75138

**KEYWORDS**

nonalcoholic fatty liver disease; NAFLD; nonalcoholic steatohepatitis; NASH; gastroenterology; elastography

## Introduction

Nonalcoholic fatty liver disease (NAFLD) is the most prevalent cause of chronic liver disease that encompasses a broad spectrum of clinical liver pathologies, including simple steatosis and nonalcoholic steatohepatitis (NASH), which could advance to fibrosis and cirrhosis, and may lead to liver failure or hepatocellular carcinoma (HCC) [1]. The worldwide prevalence of NAFLD is estimated to be 24% to 30% [2,3]. However, estimates for India vary and range from 9% to 53%. In India, NAFLD and NASH have been identified as predominant causes of cirrhosis, HCC, and liver transplantation [4]. A total of 20% of patients with NASH were reported to develop cirrhosis, and 30% to 40% of the patients with NASH cirrhosis died from liver-related causes [5,6].

In June 2023, an international consensus panel recommended revising the nomenclature of NAFLD and NASH to *metabolic dysfunction–associated steatotic liver disease (MASLD)* and *metabolic dysfunction–associated steatohepatitis (MASH)* to better reflect the underlying metabolic basis of the diseases [7,8]. However, the Indian National Association for Study of the Liver (INASL; 2023) continues to recommend the use of NAFLD or NASH terminology in India until broader consensus and implementation are achieved [9]. Importantly, INASL cautions that the new nomenclature "may create confusion not only among hepatologists but also among physicians and non-hepatologists who manage these patients or conduct research in this area."

Various comorbidities have been reported among patients with NAFLD and NASH, such as diabetes, obesity, and hypertension [10]. NASH affects up to 30% of patients with obesity, which poses a serious risk to their liver and cardiovascular health [11]. Although various diagnostic strategies are available for the diagnosis of NAFLD and NASH in hepatology and gastroenterology practices, it is estimated that fewer than 20% of patients with NAFLD have been diagnosed [12,13]. In addition, among the diagnosed patients, very few are referred to hepatologists and gastroenterologists [14]. The mainstay of NAFLD treatment is lifestyle changes, including weight loss and physical exercise; however, many patients with NAFLD fail to lose weight [15,16].

The growing prevalence and limited number of approved medications for NAFLD and NASH remain a challenge in the management of these conditions [17]. In addition, long-term data on improvement in fibrosis with the existing therapies are lacking [18]. Therefore, it becomes imperative to understand the challenges faced by hepatologists and gastroenterologists in the diagnosis and management of NAFLD.

In a recent global survey, 68% of physicians reported lack of awareness as the most common reason for the limited screening of NAFLD, which can worsen the burden of NAFLD in the future [19]. The major barriers to NAFLD management include the lack of Indian data on the management of patients with NAFLD, low screening rates of vulnerable populations, and the paucity of suitable treatment options [9,17].

Therefore, to understand real-world perspectives of Indian hepatologists and gastroenterologists on NAFLD and NASH, a cross-sectional questionnaire-based disease perspective (DRIVE) survey was carried out between May 2023 and July 2023. As this survey was conducted *before* the global shift to MASLD and MASH terminology (June 2023), the survey questionnaire continued to use the NAFLD and NASH terms; therefore, these terms are retained throughout this paper for accuracy and consistency in reporting survey data.

This survey sought to capture the current clinical practices and perspectives on NAFLD and NASH management across India. The increasing global burden of metabolic liver disease and the variability in health care infrastructure and clinical practices across countries present significant challenges to standardized care. Evaluating diagnostic and therapeutic strategies from large-population settings such as India—where MASLD prevalence is rising rapidly—can provide critical insights, which can offer valuable lessons to other low- and middle-income countries facing similar implementation barriers [7,8].

## Methods

### Participants and Procedures

This web-based, cross-sectional survey was conducted online between May 15, 2023, and July 31, 2023, among practicing gastroenterologists and hepatologists across India to understand their perspectives on NAFLD and NASH. Eligible participants included practicing gastroenterologists and hepatologists working in various health care setups such as multispecialty hospitals, clinics, nursing homes, and government hospitals or medical colleges across the country who voluntarily agreed to participate in this survey. Invitations to participate were distributed electronically. Details on the number of invitations distributed and the source of the contact list were not recorded. The online survey questionnaire link was distributed electronically (via email), and gastroenterologists who provided a full response to the survey were included (N=609) in the analysis.

### Survey Questionnaire

A predefined, structured, and self-administered questionnaire was developed by the study investigators based on national (INASL) [9] and international clinical practice guidelines for NAFLD and NASH management [7,20]. The questionnaire included 34 questions distributed across 3 domains: disease perspectives (16 items), diagnostic modalities (4 items), and management strategies (14 items). Items were reviewed by hepatology experts to ensure relevance and clarity. Both closed-ended questions with predefined response options and open-ended questions for additional comments were included.



XSL•FO
RenderX

The "Disease Perspectives" section included 16 questions on the monthly NAFLD caseload, the percentage of patients with NAFLD who had NASH, biopsy rates among patients with NASH, common fibrosis stages at presentation, the likelihood of progression to cirrhosis within 5 years, and actions taken upon incidental NAFLD detection. It also covered risk factors, comorbidities, and symptoms commonly associated with NAFLD.

The "Diagnostic Modalities" section encompassed 4 questions focused on recommended tests for NAFLD diagnosis, additional laboratory data considered, scoring systems used, and indications for liver biopsies.

Finally, 14 questions in the "Management Strategies" section included referral patterns to dietitians, commonly prescribed medications, follow-up frequencies, monitoring methods, guidelines followed, challenges in managing NASH, and anticipated future treatment options. The pharmacotherapy domain of the questionnaire included the following response options: vitamin E or antioxidant formulations, saroglitazar, pioglitazone, and a separate category labelled "hepatoprotective agents," which referred to ursodeoxycholic acid (UDCA), silymarin, and related antioxidant or herbal formulations commonly used in India [14,19].

The questionnaire included both open- and closed-ended questions to capture detailed insights. Demographic data on participants' practice settings, clinical experience, the number of patients with NAFLD consulted within a given time frame, and specialty referral patterns for patients with NAFLD were also collected (Multimedia Appendix 1).

### Reporting Standards

This survey is reported in accordance with the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) guidelines (Multimedia Appendix 2).

### Statistical Analysis

Survey data were analyzed using descriptive statistics, with results summarized as frequencies and percentages to illustrate the distribution of responses across different categories. Responses to open-ended questions were also summarized descriptively to identify common themes. Summary statistics were used for numerical variables wherever applicable. All analyses were conducted using SPSS software (version 16.0; IBM Corp).

### Ethical Considerations

This survey was reviewed and approved by the Royal Pune Independent Ethics Committee, Pune, Maharashtra, India (RPIEC110123; dated January 10, 2023). This was a physician-based survey with no involvement of patients or collection of patient-identifiable data. Hence, patient informed consent was not applicable.

Participation in this survey was voluntary and anonymous. Before accessing the survey, each participant was presented with an electronic informed consent form describing the survey objectives, the confidentiality of responses, the voluntary nature of participation, and the right to withdraw before final submission. Only those who provided electronic consent could proceed to complete the questionnaire. No personal identifiers were collected, and data were analyzed in aggregate.

No monetary or nonmonetary compensation was provided to the participants.

## Results

### Survey Participants' Demographics

A total of 609 physicians (gastroenterologists: n=556, 91.3%; hepatologists: n=53, 8.7%) participated in this survey from various centers across India. All the participating physicians had completed the survey in full. Most of the participating physicians were male individuals (n=574, 94.3%). In this survey, most physicians (n=213, 35%) had 6 to 10 years of experience in clinical practice. Overall, 341 (56%) physicians predominantly practiced in multispecialty hospitals, while 120 (19.7%) worked in mixed settings (Table 1).



XSL•FO
RenderX

**Table 1.** Demographic details of the participants (N=609).

| Variable | Participants, n (%) |
|---|---|
| **Sex** | |
| Male | 574 (94.3) |
| Female | 35 (5.7) |
| Intersex | 0 (0) |
| **Duration of clinical practice (y)** | |
| <5 | 149 (24.5) |
| 6-10 | 213 (35) |
| 11-20 | 178 (29.2) |
| >20 | 69 (11.3) |
| **Predominant type of practice** | |
| Multispecialty hospital | 341 (56) |
| Mixed | 120 (19.7) |
| Clinic or nursing home | 95 (15.6) |
| Government hospital or medical college | 53 (8.7) |

## Disease Perspective

Of the 609 physicians, 336 (55.2%) reported that the proportion of patients with NAFLD out of total patients examined each month was 25% to 50%. Furthermore, 220 (36.1%) physicians responded that 10% to 20% of their patients with NAFLD had NASH, and 523 (85.9%) physicians reported that less than 5% of their patients with NASH underwent biopsy. Patients with NAFLD presented to their physicians at varying stages of fibrosis. Most physicians (n=319, 52.4%) reported that majority of their patients with NAFLD first presented at stage 2 of fibrosis, while 204 (33.5%) physicians reported that 5% to 10% of patients with NASH progressed to cirrhosis over the course of 5 years.

Consulting physicians (n=491, 80.6%) were the most common specialties that referred patients with NAFLD to gastroenterologists and hepatologists (Table 2). According to 227 (37.3%) of the physicians, 25% to 50% of their patients with NAFLD were diagnosed incidentally. Among the patients with incidentally diagnosed NAFLD, various actions undertaken by physicians included investigation for the presence of metabolic alterations and the absence of other liver diseases (n=426, 70%), assessment for cardiovascular risk and liver disease (n=346, 56.8%), assessment for the risk of metabolic alterations and liver function with a noninvasive scoring system (n=499, 81.9%), investigation for presence of alcohol consumption (n=394, 64.7%), advice on lifestyle modification and follow-up (n=487, 80%), and initiation of pharmacotherapy (n=145, 23.8%), as reported by physicians.

Most physicians considered obesity (n=583, 95.7%) and diabetes mellitus (n=579, 95.1%) as the risk factors for NAFLD (Table 3). A total of 295 (48.4%), 149 (24.5%), and 300 (49.3%) physicians reported that 20% to 40% of their patients with NAFLD had diabetes, obesity, and dyslipidemia, respectively.

Of the 609 physicians, 331 (54.3%) reported that 10% to 30% of their patients with NAFLD required a multispecialty management approach. The proportion of lean NASH diagnosed among all patients with NASH per year was 3% to 5%, as reported by 224 (36.8%) physicians. Approximately 370 (60.8%) physicians recommended screening the family members of patients with NAFLD. Most physicians (n=345, 56.7%) reported that fewer than 10% of their patients with NAFLD were symptomatic. The most common symptoms reported to physicians by their patients with NAFLD were right upper quadrant pain (n=425, 69.8%), fatigue or tiredness (n=421, 69.1%), a feeling of fullness (n=286, 47%), and bloating (n=270, 44.3%; Table 4).

**Table 2.** Predominant specialties referring patients with nonalcoholic fatty liver disease to gastroenterologists or hepatologists (N=609).

| Referring specialties | Survey responses, n (%) |
|---|---|
| Consulting physician | 491 (80.6) |
| Endocrinologist | 85 (14) |
| Cardiologist | 16 (2.6) |
| General practitioner | 7 (1.1) |
| Surgeons | 3 (0.5) |
| Gynecologists | 2 (0.3) |
| Others | 5 (0.8) |



XSL•FO
RenderX

**Table 3.** Risk factors for nonalcoholic fatty liver disease (NAFLD) as reported by survey participants.

| Risk factor | Respondents, n (%) |
|---|---|
| Obesity | 583 (95.7) |
| Diabetes | 579 (95.1) |
| Dyslipidemia | 547 (89.8) |
| Hypothyroidism | 381 (62.6) |
| Polycystic ovary syndrome | 367 (60.3) |
| Hypertension | 335 (55) |
| Sleep apnea | 281 (46.1) |
| Ischemic heart disease | 189 (31) |

**Table 4.** Symptoms commonly reported by patients with nonalcoholic fatty liver disease (NAFLD) as reported by survey participants (N=609).

| Symptoms commonly reported by patients with NAFLD | Respondents, n (%) |
|---|---|
| Right upper quadrant pain | 425 (69.8) |
| Fatigue or tiredness | 421 (69.1) |
| A feeling of fullness | 286 (47) |
| Bloating | 270 (44.3) |
| Weakness | 184 (30.2) |
| Asymptomatic | 4 (0.7) |
| Other symptoms[a] | 9 (1.5) |

[a]Other symptoms include gastroesophageal reflux disease, incidental detection, incomplete evacuation of bowel, lethargy, loss of appetite, low mood, poor appetite, sense of indigestion, discomfort or pain after meals, and other complaints.

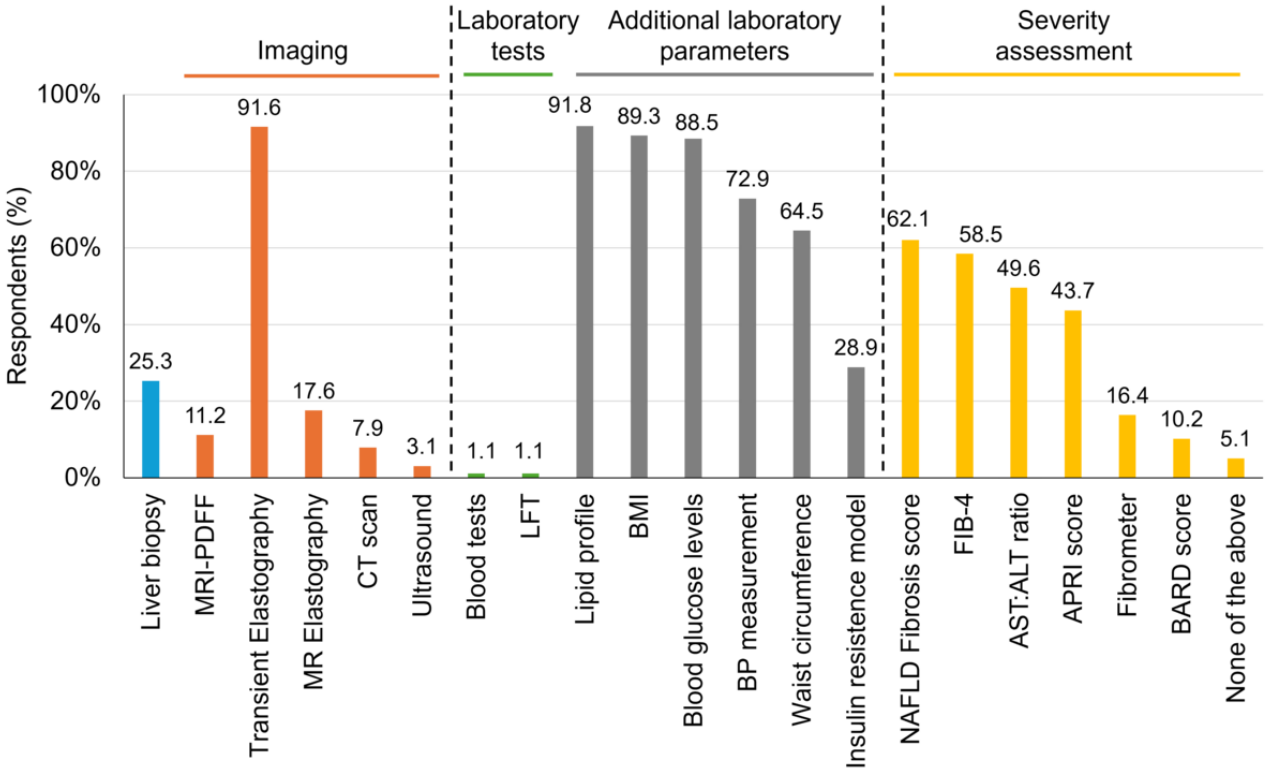
## Diagnostic Modalities

For the diagnosis of NAFLD, most physicians preferred transient elastography (TE; n=558, 91.6%) followed by liver biopsy (n=154, 25.3%), magnetic resonance elastography (n=107, 17.6%), and magnetic resonance imaging–proton density fat fraction (n=68, 11.2%), computed tomography scan (n=48, 7.9%), and ultrasound (n=19, 3.1%). Additional tests or assessments routinely advised in patients diagnosed with NAFLD were lipid profile, BMI, blood glucose levels, and so on. To assess the severity of NAFLD, scores such as the NAFLD fibrosis score, Fibrosis-4 (FIB-4) score, aspartate aminotransferase-to-alanine aminotransferase ratio, and aspartate aminotransferase-to-platelet ratio index score were used (Figure 1). A total of 286 (47%) of the physicians reported that the main intent of advising a biopsy in patients with NAFLD was to determine the stage of fibrosis for prognostic purposes, while 236 (38.8%) of the physicians advised biopsy to differentiate NASH from steatosis for diagnostic purposes.



XSL•FO
RenderX

**Figure 1.** Summary of the clinicians' preference for various diagnostic modalities to diagnose and differentiate between nonalcoholic fatty liver disease (NAFLD) and its subtypes in patients. ALT: alanine aminotransferase; APRI: aspartate aminotransferase-to-platelet ratio index; AST: aspartate aminotransferase; BARD score: BMI, AST:ALT ratio, and diabetes mellitus; BP: blood pressure; CT: computed tomography; FIB-4: Fibrosis-4; HOMA-IR: homeostatic model assessment of insulin resistance; LFT: liver function test; MR: magnetic resonance; MRI-PDFF: magnetic resonance imaging–proton density fat fraction.




## Management Strategies

Of the 609 physicians, pharmacotherapy, along with dietary and lifestyle modifications, was prescribed by 414 (68%) physicians for the management of patients with NAFLD, while 195 (32%) physicians used only dietary and lifestyle modifications to manage their patients. It was reported that more than 30% of the patients were referred to a dietitian for dietary management by 278 (45.6%) physicians. A total of 499 (81.9%) physicians allowed no alcohol consumption for patients with NASH, and 110 (18.1%) physicians allowed 10 to 30 g of alcohol per day.

Antioxidant vitamins were the most often prescribed drugs (n=543, 89.2%) for the treatment of NAFLD and NASH. Other commonly prescribed drugs were saroglitazar (n=522, 85.7%), hepatoprotective agents (n=366, 60.1%), obeticholic acid (n=255, 41.9%), and pioglitazone (n=141, 23.2%; Table 5).

In this survey, of the 609 physicians, 525 (86.2%) reported prescribing vitamin E for treating NASH, and UDCA was prescribed by 341 (56%) physicians, while 26 (4.3%) physicians recommended alternative medicine or herbal preparations. Aside from pharmacotherapy, dietary and lifestyle modifications were also recommended to more than 30% of the patients with NAFLD.

Of the 609 physicians, 498 (81.8%) reported that they followed up with their patients with NAFLD every 3 months. To monitor the response of patients to pharmacotherapy, 558 (91.6%) physicians reported using liver function tests, while 477 (78.3%) preferred elastography, 234 (38.4%) used different scoring methods based on serum markers, and 188 (30.9%) used ultrasonography (Table 6).

Of the 609 physicians, most (n=347, 57%) referred to the American Association for Study of Liver Diseases (AASLD) guidelines, while 148 (24.3%) physicians referred to the Asian Pacific Association for the Study of the Liver (APASL) guidelines for diagnosing and managing patients with NAFLD. In terms of treatment barriers, most physicians (n=466, 76.5%) considered lack of patient awareness as a major barrier to managing NASH. Additionally, 303 (49.8%) physicians reported a lack of availability of effective drugs, and 264 (43.3%) reported the cost of evaluation and treatment as barriers (Table 7).

The drugs considered by physicians as potential future therapeutic options for the management of NAFLD and NASH included peroxisome proliferator–activated receptor α and γ agonists (n=348, 57.1%), glucagon-like peptide-1 (GLP-1) agonists (n=271, 44.5%), farnesoid X receptor agonists (n=262, 43%), and sodium-glucose cotransporter-2 inhibitors (n=219, 36%).



XSL•FO
RenderX

**Table 5.** Drugs commonly prescribed for the management of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH; N=609).

| Commonly prescribed drugs for NAFLD and NASH management | Respondents, n (%) |
|---|---|
| Antioxidant vitamins | 543 (89.2) |
| Saroglitazar | 522 (85.7) |
| Hepatoprotective agents | 366 (60.1) |
| Obeticholic acid | 255 (41.9) |
| Pioglitazone | 141 (23.2) |

**Table 6.** Investigations preferred by physicians to monitor patients' responses to pharmacotherapy (N=609).

| Clinical investigations | Survey responses, n (%) |
|---|---|
| Liver function tests | 558 (91.6) |
| Elastography | 477 (78.3) |
| Different scoring based on serum markers (aspartate aminotransferase-to-platelet ratio index score and Fibrosis-4 score) | 234 (38.4) |
| Ultrasonography | 188 (30.9) |
| Liver biopsy | 40 (6.6) |
| Magnetic resonance imaging | 29 (4.8) |

**Table 7.** Barriers to the management of nonalcoholic steatohepatitis (NASH; N=609).

| Barriers in NASH management | Respondents, n (%) |
|---|---|
| Lack of awareness in patients | 466 (76.5) |
| Lack of availability of effective drugs | 303 (49.8) |
| Cost of evaluation and treatment | 264 (43.3) |
| Time constraint | 105 (17.2) |
| No barriers | 9 (1.5) |

# Discussion

## Principal Findings

This nationwide survey of 609 Indian gastroenterologists and hepatologists provides a comprehensive understanding of real-world practices in the diagnosis and management of NAFLD and NASH. The findings highlight that NAFLD accounts for a substantial proportion of liver disease cases seen in daily practice, with obesity and diabetes being the most frequently coexisting metabolic comorbidities, consistent with the well-documented pathophysiologic overlap between metabolic syndrome and NAFLD in Indian patients [1,2,4]. Notably, this survey was conducted between May 2023 and July 2023, before the official nomenclature change from NAFLD and NASH to MASLD and MASH was announced globally; therefore, the previous terminology was used throughout the questionnaire and data reporting to ensure consistency.

Although TE was the most frequently used diagnostic modality in our survey (reported by 558/609, 91.6% of the physicians), its diagnostic value lies primarily in the noninvasive assessment of liver fibrosis rather than the differentiation between simple steatosis and steatohepatitis (MASH). Current international guidelines from AASLD, European Association for the Study of the Liver (EASL), and APASL recommend an initial screening with simple serum-based fibrosis indices such as FIB-4 or the NAFLD fibrosis score, followed by TE as a secondary assessment in patients at intermediate or high risk [9,20,21]. The predominant reliance on TE observed in our cohort likely reflects its availability and routine use in tertiary centers, where most patients are referred cases from consulting physicians or secondary-care facilities. In such settings, TE serves as a valuable tool for fibrosis staging and risk stratification. However, optimizing its use in accordance with global stepwise algorithms—including initial screening with low-cost serum indices—would ensure more efficient resource use and enhance adherence to international best practices across all levels of care in India.

In contrast, pharmacological management patterns were dominated by saroglitazar (n=522, 85.7%) and vitamin E (n=525, 86.2%), and hepatoprotective agents such as UDCA and silymarin remained frequently prescribed despite limited supporting evidence [17,22]. These findings illustrate the diversity of real-world clinical approaches in India and highlight opportunities to further align everyday practice with evolving evidence and guideline-based recommendations.



XSL•FO
RenderX

### Comparison With Prior Work

Our findings are consistent with previously published Indian studies and surveys showing variability in NAFLD diagnosis and treatment pathways [3,5,17]. Similar reliance on TE and limited use of validated scoring systems was reported by Singh et al [6] and Duseja et al [9], emphasizing the need for broader adoption of noninvasive fibrosis indices for early risk stratification. Internationally, AASLD, EASL, and APASL recommend a tiered diagnostic pathway beginning with simple serum-based scores, followed by elastography or imaging only when advanced fibrosis is suspected [20,21].

The predominance of saroglitazar use (n=522, 85.7%) in this survey reflects its availability and regulatory approval for NAFLD in India. However, current global guidelines have not endorsed saroglitazar as a first-line therapy, and its approval outside India remains limited [15,17]. Although Indian real-world studies report biochemical and metabolic improvements, long-term histological and outcome data remain limited. Our survey did not capture details regarding cost, accessibility, or physician perceptions influencing prescription behavior; therefore, these findings should be interpreted as descriptive of current practice rather than indicative of comparative efficacy or cost-effectiveness. This finding underscores the need to examine economic and awareness factors shaping therapy choices in future research.

GLP-1 receptor agonists (RAs) such as semaglutide show histological improvement in NASH and metabolic benefits in global phase 3 trials [23,24]. When the DRIVE survey was conducted, injectable semaglutide was not approved in India; therefore, questions on its clinical use were not included. Instead, GLP-1 RAs were listed among potential future therapeutic options, for which only a small proportion of clinicians expressed interest. Given that semaglutide is still not approved for NASH in India and requires parenteral administration, widespread use is currently limited. Future studies should assess the real-world uptake, feasibility, and outcomes associated with GLP-1–based therapies in Indian clinical practice.

The term “hepatoprotective agents” in our questionnaire encompassed widely used drugs such as UDCA, silymarin, and antioxidant formulations. These agents are not recommended by AASLD, EASL, or INASL for NAFLD management due to the absence of histological benefit [9,21]. Their continued use by more than half of the surveyed clinicians (n=366, 60.1%) likely reflects therapeutic inertia, overreliance on empirical, perceived safety, and patient demand. However, this pattern must also be interpreted in the context of limited pharmacological options available in India during the survey period—till early 2025, saroglitazar remained the only approved agent for NASH, and no other globally validated drugs had received regulatory clearance. In the absence of alternative evidence-based therapies, clinicians may have continued to rely on hepatoprotective agents to provide symptomatic or biochemical improvement. These findings suggest opportunities to strengthen clinician education; wider dissemination of updated practice guidelines; and accelerated access to effective, evidence-based therapies to address the unmet clinical needs in NAFLD management.

Globally, there has been a paradigm shift in nomenclature from *NAFLD* and *NASH* to *MASLD* and *MASH* [7,8]. This change was formally endorsed in June 2023 by major international societies, including AASLD and EASL, reflecting a conceptual move away from alcohol exclusion criteria toward emphasizing metabolic dysfunction as the central driver of disease. The DRIVE survey was conducted before this global transition; therefore, it used NAFLD and NASH terminology, which has been retained in this study to maintain accuracy and continuity of the survey data.

In the INASL 2023 Guidance Paper, it was explicitly acknowledged that the recent nomenclature change “may create confusion not only among hepatologists but also among physicians and non-hepatologists who manage these patients or conduct research in this area” [9]. Accordingly, INASL continues to recommend the use of NAFLD and NASH terminology within India until uniform adoption of the MASLD and MASH framework is achieved.

In addition to metabolic drivers of steatotic liver disease, HCC may arise in patients with MASLD even in the absence of cirrhosis. Onzi et al [25] reported that a significant proportion of HCC cases in NAFLD occur in noncirrhotic livers, suggesting that hepatocarcinogenesis in MASLD involves metabolic and inflammatory pathways independent of advanced fibrosis. Soldera et al [26] further emphasized the need to integrate risk factors such as diabetes and obesity into HCC surveillance frameworks, highlighting the relevance of the Barcelona Clínic Liver Cancer protocol for risk-based screening. Given India’s rising burden of diabetes and obesity, these findings support reconsidering surveillance criteria to include selected patients with MASLD who are noncirrhotic at elevated risk for HCC.

### Strengths and Limitations

This survey has several notable strengths. First, it captured perspectives from a large, nationally representative cohort of 609 practicing gastroenterologists and hepatologists across diverse health care settings in India, providing a robust overview of real-world diagnostic and therapeutic approaches. Second, it used a structured, predefined questionnaire based on national and international clinical guidance, ensuring consistency and comparability of responses. Third, the survey included both hepatologists and gastroenterologists, reflecting multidisciplinary viewpoints and providing a realistic picture of current clinical practice. Fourth, because the survey was conducted immediately before the global transition from NAFLD and NASH to MASLD and MASH terminology, it provides a baseline on physician practices before the nomenclature change—information that will be valuable for future longitudinal comparisons.

However, several limitations merit mention. First, responses were self-reported and may be subject to recall or social-desirability bias. Second, as participation was voluntary and the total number of invitations sent was not recorded, representativeness and the response rate could not be assessed. Third, as this was a noncomparative, exploratory survey aimed



XSL•FO
RenderX

at describing clinical practice patterns, only descriptive statistical analyses were conducted, and inferential statistics were not performed; hence, causal inferences cannot be drawn. Fourth, this survey reflects physician perceptions and may not fully represent patient-level practice data. Fifth, newer pharmacological agents and diagnostic technologies may have been underrepresented, given the survey's cross-sectional design and timing before the MASLD nomenclature shift. Sixth, the survey did not include specific questions on HCC risk assessment or surveillance practices, an important area of growing relevance given the emerging evidence of HCC development in noncirrhotic MASLD. Finally, the survey did not include questions on drug affordability, availability, or rationale for therapy selection, which limits the interpretation of the observed preference for therapeutic options. Therefore, while the survey provides valuable insights into prevailing clinical practices, the findings should be interpreted cautiously and may not be fully generalizable to all health care settings, including primary care.

### Future Directions

Future research should explore longitudinal changes in diagnostic and treatment practices as the MASLD and MASH terminology and updated clinical frameworks become integrated into Indian hepatology practice. Strengthening implementation and adherence to existing India-specific consensus guidelines such as those issued by INASL [9] will be critical to ensure diagnostic accuracy and affordability, particularly by promoting the stepwise use of noninvasive scores (FIB-4 and NAFLD fibrosis scores) before TE. Educational interventions and continuous medical education programs can be instrumental in promoting rational pharmacotherapy and evidence-based management [18]. However, a major barrier in the Indian context remains the limited availability of approved, effective agents for NAFLD and NASH. Addressing this gap will require focused research and development on novel therapies—particularly those demonstrating antifibrotic effects—along with timely regulatory approval and improved access to such agents once available.

National liver societies should prioritize promoting evidence-based treatment practices and gradually phase out the use of hepatoprotective agents with limited clinical benefits, alongside structured dissemination of MASLD nomenclature and risk-adapted HCC surveillance even in noncirrhotic metabolic liver disease [7,26]. Establishing multicenter registries and real-world databases can further support health-economic evaluations and inform locally relevant guideline updates. Although this survey focused on Indian clinicians, its findings have wider international relevance, as the challenges identified—limited patient awareness leading to delayed diagnosis, few effective or approved treatment options, and discrepancies between drugs approved in India (eg, saroglitazar) and those endorsed internationally (eg, GLP-1 RAs)—mirror challenges faced by other low- and middle-income countries [13]. These insights can help inform global implementation strategies for MASLD management in diverse health care settings.

### Conclusions

This nationwide survey highlights considerable variability between prevailing clinical practices in India and international guidelines for the diagnosis and management of metabolic dysfunction–associated liver disease. While TE has become the most widely used noninvasive test, reliance on it without previous application of simple scores such as FIB-4 or the NAFLD fibrosis score indicates gaps in stepwise risk stratification. Saroglitazar remains the most commonly prescribed drug owing to its domestic approval and availability, but its long-term histological efficacy and cost-effectiveness require further evidence. The continued use of hepatoprotective agents such as UDCA and silymarin highlights the persistence of non–evidence-based treatments.

The transition from NAFLD and NASH to MASLD and MASH nomenclature provides an opportunity for India to harmonize diagnostic frameworks, enhance physician education, and modernize treatment algorithms. Future national efforts should focus on stepwise noninvasive staging, rational pharmacotherapy, risk-adapted surveillance for HCC—including in selected patients with MASLD who are noncirrhotic—and large-scale real-world registries to monitor outcomes. Harmonizing Indian clinical practice with evolving international standards is imperative to enhance diagnostic accuracy, consistency, and patient outcomes in metabolic dysfunction–associated liver disease.

## Acknowledgments

The authors thank all the gastroenterologists and hepatologists who participated in this survey. Authors acknowledge the efforts and contributions of the doctors involved in the survey. The authors thank THB c/o Sekhmet Technologies Private Limited for the overall administration and management of the project as a Contract Research Organization (CRO). The authors also thank THB c/o Sekhmet Technologies Private Limited for contributing to manuscript writing and editing. Anjaly Krishnan from THB c/o Sekhmet Technologies Private Limited is responsible for statistical analysis, data analysis, and data visualization.

## Funding

This survey was funded by Sun Pharma Laboratories Limited.

## Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.



XSL•FO
RenderX

## Conflicts of Interest

Authors SS, SD, CK, NM, AM, and SCM are full-time employees of Sun Pharma Laboratories Limited. Author YG was a full-time employee of Sun Pharma Laboratories Limited. Authors AA and AK have no conflicts of interest.

## Authors' Contributions

Conceptualization: YG, SS, CK, NM, AM, SCM, AA, AK;
Data Curation: AA, AK;
Formal Analysis: YG, CK, NM, AM, SCM;
Funding acquisition: AA, AK;
Investigation: AA, AK;
Methodology: YG, CK, NM, AM, SCM;
Project administration: YG, CK, NM, AM, SCM, AA, AK, SS, SD;
Resources: AA, AK;
Software: AA, AK;
Supervision: AA, YG, SS, SD, CK, NM, AM, SCM, AK;
Validation: AA, YG, SS, SD, CK, NM, AM, SCM, AK;
Visualization: AA, YG, SS, SD, CK, NM, AM, SCM, AK;
Writing – Original draft: SS, AA, SD, CK, NM, AM, SCM, AK;
Writing – reviewing & editing: AA, YG, SS, SD, CK, NM, AM, SCM, AK;
All authors read and approved the final manuscript.

Multimedia Appendix 1
Survey questionnaire form administered to the gastroenterologists and hepatologists.
[DOCX File , 31 KB - ijmr_v15i1e75138_app1.docx ]

Multimedia Appendix 2
CHERRIES guidelines.
[DOCX File , 23 KB - ijmr_v15i1e75138_app2.docx ]

## References

1. Cobbina E, Akhlaghi F. Non-alcoholic fatty liver disease (NAFLD) - pathogenesis, classification, and effect on drug metabolizing enzymes and transporters. Drug Metab Rev 2017 May;49(2):197-211 [FREE Full text] [doi: 10.1080/03602532.2017.1293683] [Medline: 28303724]
2. Sherif ZA. The rise in the prevalence of nonalcoholic fatty liver disease and hepatocellular carcinoma. In: Gad EH, editor. Nonalcoholic Fatty Liver Disease - An Update. New York, NY: IntechOpen; 2019.
3. Younossi ZM, Golabi P, de Avila L, Paik JM, Srishord M, Fukui N, et al. The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: a systematic review and meta-analysis. J Hepatol 2019 Oct;71(4):793-801. [doi: 10.1016/j.jhep.2019.06.021] [Medline: 31279902]
4. De A, Duseja A. Nonalcoholic fatty liver disease: Indian perspective. Clin Liver Dis (Hoboken) 2021 Sep;18(3):158-163 [FREE Full text] [doi: 10.1002/cld.1141] [Medline: 34691404]
5. Armandi A, Bugianesi E. Natural history of NASH. Liver Int 2021 Jun;41 Suppl 1(Suppl 1):78-82 [FREE Full text] [doi: 10.1111/liv.14910] [Medline: 34155792]
6. Singh S, Allen AM, Wang Z, Prokop LJ, Murad MH, Loomba R. Fibrosis progression in nonalcoholic fatty liver vs nonalcoholic steatohepatitis: a systematic review and meta-analysis of paired-biopsy studies. Clin Gastroenterol Hepatol 2015 Apr;13(4):643-54.e1 [FREE Full text] [doi: 10.1016/j.cgh.2014.04.014] [Medline: 24768810]
7. Rinella ME, Lazarus JV, Ratziu V, Francque SM, Sanyal AJ, Kanwal F, NAFLD Nomenclature consensus group. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. J Hepatol 2023 Dec;79(6):1542-1556 [FREE Full text] [doi: 10.1016/j.jhep.2023.06.003] [Medline: 37364790]
8. Eslam M, Newsome PN, Sarin SK, Anstee QM, Targher G, Romero-Gomez M, et al. A new definition for metabolic dysfunction-associated fatty liver disease: an international expert consensus statement. J Hepatol 2020 Jul;73(1):202-209 [FREE Full text] [doi: 10.1016/j.jhep.2020.03.039] [Medline: 32278004]
9. Duseja A, Singh SP, De A, Madan K, Rao PN, Shukla A, Shalimar, et al. Indian National Association for Study of the Liver (INASL) guidance paper on nomenclature, diagnosis and treatment of nonalcoholic fatty liver disease (NAFLD). J Clin Exp Hepatol 2023;13(2):273-302 [FREE Full text] [doi: 10.1016/j.jceh.2022.11.014] [Medline: 36950481]
10. Huh Y, Cho YJ, Nam GE. Recent epidemiology and risk factors of nonalcoholic fatty liver disease. J Obes Metab Syndr 2022 Mar 30;31(1):17-27 [FREE Full text] [doi: 10.7570/jomes22021] [Medline: 35332111]



XSL•FO
RenderX

11. Billeter AT, Reiners B, Seide SE, Probst P, Kalkum E, Rupp C, et al. Comparative effectiveness of medical treatment metabolic surgery for histologically proven non-alcoholic steatohepatitis and fibrosis: a matched network meta-analysis. Hepatobiliary Surg Nutr 2022 Oct;11(5):696-708 [FREE Full text] [doi: 10.21037/hbsn-21-5] [Medline: 36268239]
12. Alexander M, Loomis AK, Fairburn-Beech J, van der Lei J, Duarte-Salles T, Prieto-Alhambra D, et al. Real-world data reveal a diagnostic gap in non-alcoholic fatty liver disease. BMC Med 2018 Aug 13;16(1):130 [FREE Full text] [doi: 10.1186/s12916-018-1103-x] [Medline: 30099968]
13. Younossi ZM, Ong JP, Takahashi H, Yilmaz Y, Eguc Hi Y, El Kassas M, Global Nonalcoholic Steatohepatitis Council. A global survey of physicians knowledge about nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol 2022 Jun;20(6):e1456-e1468 [FREE Full text] [doi: 10.1016/j.cgh.2021.06.048] [Medline: 34229038]
14. Said A, Gagovic V, Malecki K, Givens ML, Nieto FJ. Primary care practitioners survey of non-alcoholic fatty liver disease. Ann Hepatol 2013;12(5):758-765 [FREE Full text] [Medline: 24018493]
15. Marra F, Lotersztajn S. Pathophysiology of NASH: perspectives for a targeted treatment. Curr Pharm Des 2013;19(29):5250-5269 [FREE Full text] [doi: 10.2174/1381612811319990344] [Medline: 23394092]
16. Takahashi Y, Sugimoto K, Inui H, Fukusato T. Current pharmacological therapies for nonalcoholic fatty liver disease/nonalcoholic steatohepatitis. World J Gastroenterol 2015 Apr 07;21(13):3777-3785 [FREE Full text] [doi: 10.3748/wjg.v21.i13.3777] [Medline: 25852263]
17. Sharma M, Premkumar M, Kulkarni AV, Kumar P, Reddy DN, Rao NP. Drugs for non-alcoholic steatohepatitis (NASH): quest for the Holy Grail. J Clin Transl Hepatol 2021 Feb 28;9(1):40-50 [FREE Full text] [doi: 10.14218/JCTH.2020.00055] [Medline: 33604254]
18. Friedman SL, Pinzani M. Hepatic fibrosis 2022: unmet needs and a blueprint for the future. Hepatology 2022 Feb;75(2):473-488. [doi: 10.1002/hep.32285] [Medline: 34923653]
19. Driessen S, de Jong VD, van Son KC, Klompenhouwer T, Colardelle Y, Alings M, et al. A global survey of health care workers' awareness of non-alcoholic fatty liver disease: the AwareNASH survey. United European Gastroenterol J 2023 Sep;11(7):654-662 [FREE Full text] [doi: 10.1002/ueg2.12445] [Medline: 37563849]
20. European Association for the Study of the Liver. EASL clinical practice guidelines on non-invasive tests for evaluation of liver disease severity and prognosis - 2021 update. J Hepatol 2021 Sep;75(3):659-689. [doi: 10.1016/j.jhep.2021.05.025] [Medline: 34166721]
21. Rinella ME, Neuschwander-Tetri BA, Siddiqui MS, Abdelmalek MF, Caldwell S, Barb D, et al. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. Hepatology 2023 May 01;77(5):1797-1835 [FREE Full text] [doi: 10.1097/HEP.0000000000000323] [Medline: 36727674]
22. Xiang Z, Chen YP, Ma KF, Ye YF, Zheng L, Yang YD, et al. The role of ursodeoxycholic acid in non-alcoholic steatohepatitis: a systematic review. BMC Gastroenterol 2013 Sep 23;13:140 [FREE Full text] [doi: 10.1186/1471-230X-13-140] [Medline: 24053454]
23. Newsome PN, Buchholtz K, Cusi K, Linder M, Okanoue T, Ratziu V, NN9931-4296 Investigators. A placebo-controlled trial of subcutaneous semaglutide in nonalcoholic steatohepatitis. N Engl J Med 2021 Mar 25;384(12):1113-1124. [doi: 10.1056/NEJMoa2028395] [Medline: 33185364]
24. Younossi ZM, Loomba R, Rinella ME, Bugianesi E, Marchesini G, Neuschwander-Tetri BA, et al. Current and future therapeutic regimens for nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. Hepatology 2018 Jul;68(1):361-371 [FREE Full text] [doi: 10.1002/hep.29724] [Medline: 29222911]
25. Onzi G, Moretti F, Balbinot SS, Balbinot RA, Soldera J. Hepatocellular carcinoma in non-alcoholic fatty liver disease with and without cirrhosis. Hepatoma Res 2019 Mar 11;5:7. [doi: 10.20517/2394-5079.2018.114]
26. Soldera J, Balbinot SS, Balbinot RA, Cavalcanti AG. Diagnostic and therapeutic approaches to hepatocellular carcinoma: understanding the Barcelona Clínic Liver Cancer Protocol. Clin Med Insights Gastroenterol 2016;9:67-71 [FREE Full text] [doi: 10.4137/CGast.S30190] [Medline: 27812296]


## Abbreviations

**AASLD:** American Association for Study of Liver Diseases
**APASL:** Asian Pacific Association for the Study of the Liver
**CHERRIES:** Checklist for Reporting Results of Internet E-Surveys
**DRIVE:** disease perspective
**EASL:** European Association for the Study of the Liver
**FIB-4:** Fibrosis-4
**GLP-1:** glucagon-like peptide-1
**HCC:** hepatocellular carcinoma
**INASL:** Indian National Association for Study of the Liver
**MASH:** metabolic dysfunction–associated steatohepatitis
**MASLD:** metabolic dysfunction–associated steatotic liver disease
**NAFLD:** nonalcoholic fatty liver disease



XSL•FO
RenderX

**NASH:** nonalcoholic steatohepatitis
**RA:** receptor agonist
**TE:** transient elastography
**UDCA:** ursodeoxycholic acid

*Edited by N Cahill, A Schwartz, M Balcarras; submitted 28.Mar.2025; peer-reviewed by J Soldera, C Sun, H Lin, D Poddighe; comments to author 16.Sep.2025; revised version received 11.Nov.2025; accepted 19.Dec.2025; published 02.Mar.2026.*

*Please cite as:*
*Arora A, Garje Y, Shaikh S, Dharmadhikari S, Khandhedia C, Markandeywar N, Mane A, Mehta SC, Kumar A*
*Perspectives of Indian Gastroenterologists and Hepatologists on Nonalcoholic Fatty Liver Disease Diagnosis and Management: Insights From the Nationwide Web-Based Cross-Sectional DRIVE Survey*
*Interact J Med Res 2026;15:e75138*
*URL: https://www.i-jmr.org/2026/1/e75138*
*doi:10.2196/75138*
*PMID:*

©Anil Arora, Yogesh Garje, Shagupta Shaikh, Shruti Dharmadhikari, Chintan Khandhedia, Neeraj Markandeywar, Amey Mane, Suyog C Mehta, Ashish Kumar. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 02.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Trends, Predictors, and Outcomes of Monitored Acute Care Unit Admissions in Older Adults: 10-Year Retrospective Analysis

Ludovic van Delden[1], MD; Stephan von Düring[2,3], MD, MSc; Frederic Rouyer[3], MD; Anne Rossel[4], MD; Christophe Marti[4,5], MD; Sara Cereghetti[1], MD; Xavier Roux[1,6], MD; Thibaut Desmettre[2,3], MD, PhD; Olivier Grosgurin[2,3,4,5], MD; Christophe A Fehlmann[2,3], MD

[1]Division of Critical Care Medicine, Department of Acute Medicine, Geneva University Hospitals, Geneva, Switzerland

[2]Department of Anaesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, Faculty of Medicine, University of Geneva, Geneva, Switzerland

[3]Division of Emergency Medicine, Department of Acute Medicine, Geneva University Hospitals, Rue Gabrielle Perret Gentil 4, Geneva, Switzerland

[4]Division of Internal and General Medicine, Department of Internal Medicine, Geneva University Hospitals, Geneva, Switzerland

[5]Department of Medicine, Faculty of Medicine, University of Geneva, Geneva, Switzerland

[6]Division of Internal Medicine for the Elderly, Department of Rehabilitation and Geriatrics, Geneva University Hospitals, Geneva, Switzerland

**Corresponding Author:**
Christophe A Fehlmann, MD
Department of Anaesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, Faculty of Medicine, University of Geneva, Geneva, Switzerland

## Abstract

**Background:** Global population aging places an increasing burden on health care systems. This is driven by multimorbidity, frailty, and polypharmacy. Older adults, particularly those aged 65 years or older, use emergency departments (EDs) more frequently and experience poorer outcomes. In this population, decisions regarding admission to monitored acute care units—intensive care units, intermediate care units, and operating rooms—are frequent and complex. While ED and intensive care unit use are well documented, data on monitored acute care units as a whole remain limited. Evidence on admission trends, patient characteristics, and outcomes in older adults is scarce.

**Objective:** This study aimed to describe temporal trends in monitored acute care unit admissions, identify predictors of such admissions, and assess outcomes following these admissions.

**Methods:** We conducted a retrospective cohort study using routinely collected electronic health record data. We included patients aged 65 years or older who visited the EDs of the Geneva University Hospitals, Switzerland, between 2009 and 2019. The primary outcome was admission to a monitored acute care unit. The secondary outcomes were hospital length of stay, 7-day mortality, and 1-year mortality. Logistic regression models were used to identify factors associated with monitored acute care unit admission and to assess the association between age and mortality.

**Results:** During the 10-year period, 701,838 ED visits were recorded. Annual visits increased from 56,944 to 76,368 (+34.1%). The increase was greater among patients aged 65 years or older (+56.1%) than among younger patients (+26.5%). A total of 180,189 older patients presented to the ED. Of these, 887 (0.5%) died in the ED, 97,238 (54.0%) were discharged home, 63,025 (35.0%) were admitted to a ward, and 19,039 (10.6%) were admitted to a monitored acute care unit. Monitored acute care unit admissions increased from 1379 (10.3%) in 2009 to 2240 (11.1%) in 2019. This represented an absolute increase of 62.4% and a relative increase of 0.8%. Predictors of monitored acute care unit admission included younger age, male sex, ambulance arrival, higher triage level, being married or in a relationship, not residing in a nursing home, and French as the primary language. Among patients admitted to a monitored acute care unit, mortality was 5.8% (1105/19,039) at 7 days and 22.3% (4251/18,039) at 1 year. Older age was associated with higher 7-day mortality (adjusted odds ratio 1.55, 95% CI 1.14 - 2.10) and 1-year mortality (adjusted odds ratio 1.28, 95% CI 1.08 - 1.51).

**Conclusions:** Admissions to monitored acute care units among older patients increased over time. These findings indicate a growing demand for high-level care in this population. Hospitals should adapt infrastructure and resource allocation to address the needs of an aging population.

***(Interact J Med Res 2026;15:e80629)*** doi:10.2196/80629

**KEYWORDS**

older patients; intensive care units; intermediate care units; critical care; hemodynamic monitoring; risk assessment



XSL•FO
RenderX

## Introduction

The global population is aging rapidly, exerting substantial pressure on health care systems [1,2]. With increasing life expectancy, older adults present with a higher burden of chronic illnesses, polypharmacy, and biological frailty. However, only limited data exist on the use of monitored acute care units within this population.

Individuals aged between 65 and 79 years exhibit the highest prevalence of comorbidity, while those aged 80 years and older have a 30% likelihood of presenting with at least one chronic condition [3]. This demographic shift places considerable strain on health care services. In the United States, studies have shown that patients aged older than 65 years account for 15% of all emergency department (ED) admissions, with these consultations increasing by 25% over the past decade [4,5]. These patients typically present with more severe emergencies, requiring extensive hospital resources [6]. Moreover, they experience poorer outcomes following admission compared to younger patients, particularly in the intensive care unit. In-hospital mortality rates are markedly higher for older patients during the postintensive phase [7,8], with the majority of deaths occurring within 3 months after intensive care unit discharge [9]. Furthermore, survivors frequently experience a substantial decline in quality of life, with age and frailty being crucial factors that affect the recovery of functional capacity [10].

These age-associated challenges necessitate careful consideration when evaluating older patients for admission to high-level care. Understanding these aspects is essential for health care providers in developing strategies that effectively address the specific needs of older patients, ensuring that the allocation of high-level care resources is both efficient and beneficial. While extensive research exists on ED use and intensive care unit admissions in older patients, there is a notable gap in data regarding the trends in monitored acute care unit admissions over time. Furthermore, research on monitored acute care unit admissions, encompassing intensive care units, intermediate care units, and operating rooms, as a whole entity, remains limited. Finally, intermediate care units have only recently been introduced in health care systems, and their impact on patient disposition and role as acute care providers remains to be determined.

This study aimed to address this gap by analyzing ED and monitored acute care unit admissions among patients aged 65 years or older over the past decade in a Swiss tertiary care institution. Our primary objective was to describe the trends in monitored acute care unit admissions during this period. Our secondary objectives were to identify predictors of monitored acute care unit admission, describe patient outcomes following monitored acute care unit admission, and explore the association between patient age and mortality at day 7 and 1 year postadmission.

## Methods

### Study Design

This single-center retrospective cohort study used routinely collected data from patients who visited the ED between 2009 and 2019. The study was approved by the institutional ethics committee of Geneva, Switzerland, and patient consent was waived based on Article 34 of the Swiss law on human research.

### Study Setting and Population

The study was conducted at Geneva University Hospitals (HUG), the largest medical complex in Switzerland. The HUG is a tertiary care institution with 2 adult EDs located at different sites. The main ED, which operates 24/7, handles approximately 80,000 visits annually and includes a stretcher bay and a walk-in sector, providing comprehensive emergency care. The second ED, opened in 2016, is situated at a geriatric hospital and handles approximately 5000 visits annually. It operates only during daytime hours and is dedicated to patients aged older than 75 years with low-severity conditions. Although this geriatric ED is equipped with a full radiological division, it has limited access to specialty consultants, sometimes requiring the transfer of patients who need urgent specialized consultation or surgical procedures. Hospitals within the HUG network share the same computer system, and patients' medical data are accessible in each hospital.

HUG provides specialized care across all medical fields, serving as a reference center for regional hospitals and often receiving patients from smaller centers. The main location offers 24-hour cardiac catheterization services and is the only trauma and stroke center in the region, providing immediate care for patients requiring intravenous thrombolysis, interventional radiology, or the management of life-threatening injuries.

The HUG's critical care infrastructure includes a medical-surgical intensive care unit and several specialized intermediate care units. These intermediate care units serve as an intermediary level of care between the intensive care unit and regular wards and include a cardiac intermediate care unit, a neurological and neurosurgical intermediate care unit (with a stroke unit), a surgical intermediate care unit, and 2 medical intermediate care units (one located at the main hospital and one at the geriatric hospital).

The HUG is funded through a mixed financing model, with approximately half of the budget provided by the State of Geneva and the remainder generated from hospital revenues (including patient care and insurance reimbursements) as well as dedicated research funding from national grants and private foundations. Health care financing in Switzerland is based on mandatory basic health insurance (LAMal), which ensures equal access to care for all residents and is funded through individual premiums, complemented by regulated cost-sharing mechanisms (eg, deductibles and co-payments) and optional supplementary insurance.

### Participants

We included all adult patients (aged ≥18 y) who visited either of HUG's EDs between January 1, 2009, and December 31,



XSL•FO
RenderX

2019. Patients were excluded if they (1) were aged younger than 65 years, (2) were transferred from another hospital or ED outside of HUG, or (3) refused the use of their data for research purposes. The age cutoff of 65 years was selected, as it is a widely recognized threshold for defining "older patients" and aligns with the retirement age in Switzerland.

### Variables and Data Sources

The primary outcome was monitored acute care unit admission, defined as an admission to an operating room (including surgical rooms, interventional radiology, or a cardiac catheterization laboratory), an intermediate care unit, or the intensive care unit. This outcome was based on patient trajectories within the institution. The secondary outcomes included hospital length of stay, mortality at day 7 after monitored acute care unit admission, and mortality at 1 year. Owing to the lack of linkage between the electronic health record and the national death registry, data were missing for patients who did not die in the hospital and were not readmitted beyond 1 year after the initial admission.

Patient variables extracted included age, sex, marital status (married or in a relationship vs single, divorced, or widowed), primary language, emergency triage level (based on the Swiss Emergency Triage Scale [11], ranging from 1 [highest priority] to 4 [lowest priority]), triage motive category, nursing home residence, and means of transportation to the ED (ambulance or nonambulance).

Data were electronically extracted from the hospital's data warehouse by a dedicated team, which contains all information routinely gathered for clinical use in the patient's electronic health record. These structured data can be retrieved for quality assessment or research purposes after approval from the research ethics board. Each care episode is assigned a unique number, facilitating linkage between different database subsets. There was no linkage with other databases.

### Study Size

The sample size was determined based on the study's objective to describe trends over a decade rather than on power calculations for statistical significance. With an expected inclusion of more than 150,000 patients, the sample size was considered sufficient to perform a multivariable logistic regression analysis without the risk of overfitting, based on established methodological recommendations regarding events per variable. Previous studies have shown that logistic regression models are unlikely to be overfitted when at least 10 events per predictor variable are available and that model performance remains robust even with fewer events per variable in large datasets with stable estimates. Given the large number of observations and outcome events in this study, the risk of overfitting was therefore considered minimal [12-14].

### Statistical Analysis

Data cleaning and statistical analysis were conducted using the integrated statistical software Stata/SE (version 18; StataCorp LLC).

First, the data were cleaned and standardized. Variables were reformatted to facilitate analysis. Outcomes were constructed based on patients' trajectories through the institution. To evaluate trends in monitored acute care unit admissions over time, we graphically computed the number of ED visits for patients aged 65 years or older and the number and type of monitored acute care unit admissions by year. The absolute and relative changes over time were reported. A linear regression model was used to evaluate trends over time and determine the statistical significance of the observed changes.

To explore predictors of monitored acute care unit admission, baseline variables were summarized using descriptive statistics for the overall population. Mean and SD (or median and IQR) were used for continuous variables, whereas frequency and proportion were used for categorical variables. Patients admitted to a monitored acute care unit were compared to those not admitted to a monitored acute care unit using standardized mean differences (SMDs). We used SMDs rather than *P* values, as these comparisons are descriptive and not inferential. An absolute SMD greater than 0.2 was considered to reflect a potentially meaningful imbalance. The association between age and monitored acute care unit admission was graphically represented using restricted cubic splines, a flexible regression method, with the number of knots determined by the Akaike information criterion and their placement based on recommended quantiles [14]. Independent predictors of monitored acute care unit admission were identified using a multivariable logistic regression model, with monitored acute care unit admission as the dependent variable. Potential predictors were selected based on previous literature or clinical relevance and included sex, age, marital status, nursing home residence, primary language, arrival by ambulance, emergency triage level, and triage motives. Relationships modeled with restricted cubic splines were presented as odds ratios (ORs), comparing older patients (75th percentile of age) to younger patients (25th percentile of age) [14]. A sensitivity analysis using age as a categorical variable was also performed.

To explore patient outcomes following monitored acute care unit admission, outcomes were summarized using descriptive statistics.

Finally, to explore the association between patient age and mortality at day 7 and 1 year following monitored acute care unit admission, restricted cubic splines were used. Subsequently, multivariable models similar to those used previously were applied.

Missing data frequency was reported for each variable. For the initial logistic regression model, a complete case analysis was performed. A subsequent analysis included variables with higher levels of missingness without imputation to assess changes in the coefficient estimates.

### Ethical Considerations

The study was conducted at HUG in accordance with the principles of Good Clinical Practice (the Declaration of Helsinki, 2002). This study was approved on July 11, 2022, by the institutional ethics committee of Geneva, Switzerland (project ID 2022 - 00987). Patient consent was waived by this committee. No compensation was offered to participants Data extraction generated deidentified data used for the analysis.



XSL•FO
RenderX

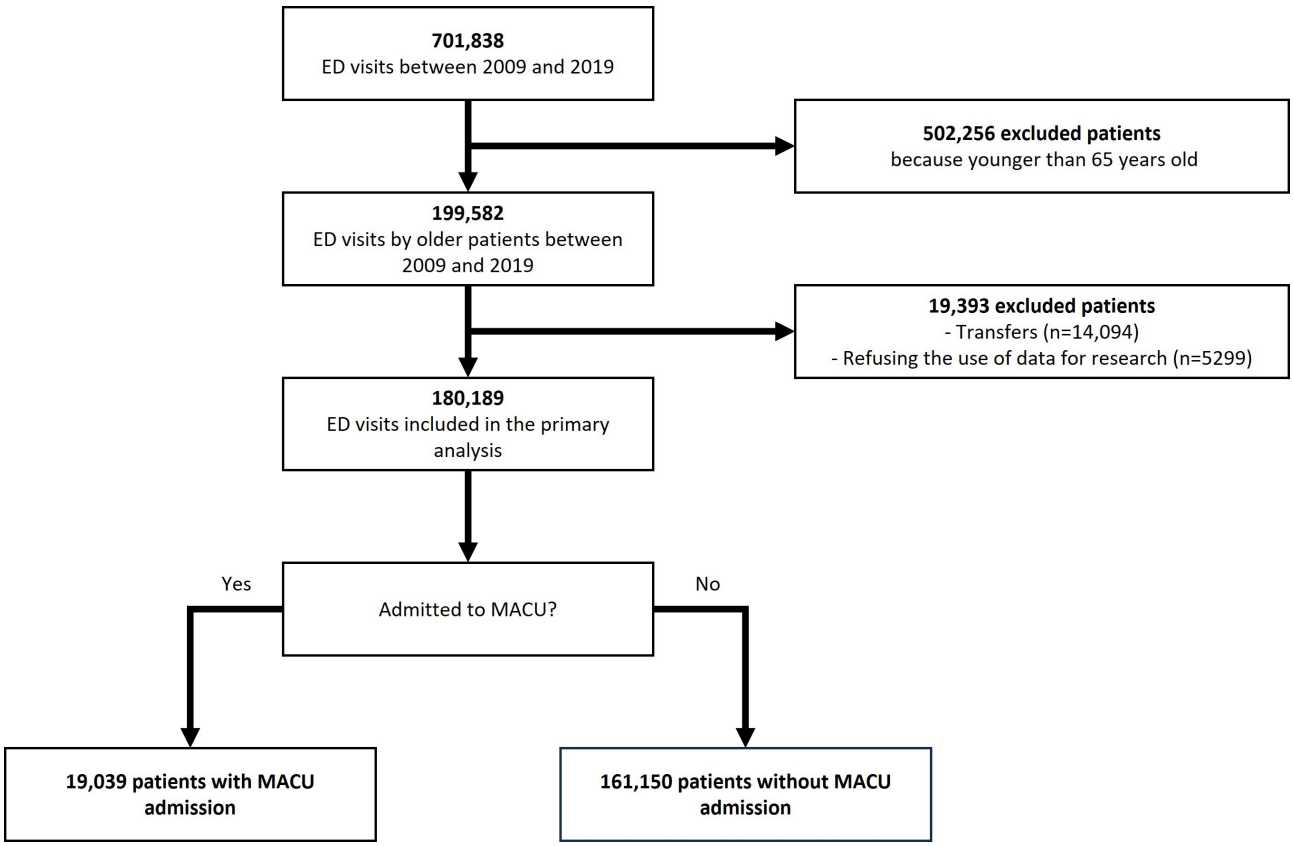
These data were subsequently anonymized for online publication.

## Results

### Inclusion of Patients and Characteristics

Between 2009 and 2019, a total of 701,838 ED visits were recorded at our institution (Multimedia Appendix 1). Of these, 180,189 (25.7%) patients met our inclusion criteria (Figure 1). The median age was 79 (IQR 72-86) years. Patient characteristics and the proportion of missing data are presented in Table 1.

**Figure 1.** Flowchart of our retrospective cohort study conducted at the Geneva University Hospitals (HUG) over a 10-y period (2009 - 2019). All adult patients (aged ≥18 y) who visited the HUG's emergency departments (EDs) were candidates for inclusion. Patients were excluded if they (1) were aged <65 y, (2) were transferred from another hospital or ED outside of HUG, or (3) refused the use of their data for research purposes. Of the 701,838 patients in our dataset, 180,189 met our inclusion criteria, representing 90.3% of all patients aged ≥65 y admitted to the ED during the study period. Of these patients, 19,039 (10.6%) were admitted to a monitored acute care unit (MACU).






XSL•FO
RenderX

**Table .** Baseline characteristics.

| | Total (N=180,189) | With MACU[a] admission (n=19,039) | Without MACU admission (n=161,150) | SMD[b] |
|---|---|---|---|---|
| Age (y), median (IQR) | 79 (72-86) | 78 (72-84) | 79 (72-86) | 0.120 |
| 65 - 74, n (%) | 61,511 (34.1) | 6890 (36.1) | 54,621 (33.9) | 0.153 |
| 75 - 84, n (%) | 66,603 (37.0) | 7668 (40.3) | 58,935 (36.6) | —[c] |
| 85 - 94, n (%) | 46,890 (26.0) | 4190 (22.0) | 42,700 (26.5) | — |
| ≥95, n (%) | 5185 (2.9) | 291 (1.5) | 4894 (3.0) | — |
| Sex, n (%) | | | | 0.150 |
| Male | 82,877 (46.0) | 10,027 (52.7) | 72,850 (45.2) | |
| Female | 97,312 (54.0) | 9012 (47.3) | 88,300 (54.8) | |
| Marital status, n (%) | | | | 0.152 |
| Married or in a relationship | 79,499 (44.1) | 9661 (50.7) | 69,838 (43.3) | |
| Single, divorced, or widowed | 100,435 (55.7) | 9297 (48.8) | 91,138 (56.6) | |
| Missing | 255 (0.1) | 81 (0.4) | 174 (0.1) | |
| Nursing home resident, n (%) | | | | 0.058 |
| No | 164,454 (91.3) | 17,552 (92.2) | 146,902 (91.2) | |
| Yes | 11,967 (6.6) | 1027 (5.4) | 10,940 (6.8) | |
| Missing | 3768 (2.1) | 460 (2.4) | 3308 (2.1) | |
| Primary language, n (%) | | | | 0.123 |
| French | 56,788 (31.5) | 6408 (34.0) | 50,308 (31.2) | |
| Italian | 5958 (3.3) | 555 (2.9) | 5403 (3.4) | |
| Spanish | 3290 (1.8) | 254 (1.3) | 3036 (1.9) | |
| German | 2069 (1.1) | 233 (1.2) | 1836 (1.1) | |
| English | 1552 (0.9) | 151 (0.8) | 1401 (0.9) | |
| Portuguese | 1355 (0.8) | 111 (0.6) | 1244 (0.8) | |
| Others | 3821 (2.1) | 326 (1.7) | 3495 (2.2) | |
| Missing | 105,536 (58.5) | 10,929 (57.4) | 64,427 (58.6) | |
| Arrival by ambulance, n (%) | | | | 0.478 |
| No | 66,881 (37.1) | 3967 (20.8) | 62,914 (39.0) | |
| Yes | 91,829 (51.0) | 13,572 (71.3) | 78,257 (48.6) | |
| Missing | 21,479 (12.0) | 1500 (7.9) | 19,979 (12.4) | |
| Triage scale, n (%) | | | | 1.152 |
| 1: vital emergency | 20,921 (11.6) | 8290 (43.5) | 12,631 (7.8) | |
| 2: urgent | 65,305 (36.2) | 7886 (41.4) | 57,419 (35.6) | |
| 3: Mildly urgent | 89,840 (49.9) | 2786 (14.6) | 87,054 (54.0) | |
| 4: nonurgent | 3788 (2.1) | 31 (0.2) | 3757 (2.3) | |
| Triage motive category, n (%) | | | | 0.759 |
| Cardiology-pneumology | 48,348 (26.8) | 7550 (39.7) | 40,798 (25.3) | |
| Neurology-psychiatry | 29,972 (16.6) | 6318 (33.2) | 23,654 (14.7) | |
| Traumatology | 30,202 (16.8) | 1598 (8.4) | 28,604 (17.8) | |



XSL•FO
RenderX

| | Total (N=180,189) | With MACU[a] admission (n=19,039) | Without MACU admission (n=161,150) | SMD[b] |
|---|---|---|---|---|
| Digestive-OB/GYN[d] | 17,731 (9.8) | 1392 (7.3) | 16,339 (10.1) | |
| Urology-nephrology | 8285 (4.6) | 263 (1.4) | 8022 (5.0) | |
| Rheumatology | 6793 (3.8) | 105 (0.6) | 6688 (4.2) | |
| Infectious disease | 5813 (3.2) | 435 (2.3) | 5378 (3.3) | |
| Dermatology | 4580 (2.5) | 134 (0.7) | 4446 (2.8) | |
| ENT[e] | 3883 (2.2) | 100 (0.5) | 3783 (2.3) | |
| Others | 24,263 (13.5) | 1107 (5.8) | 23,156 (14.4) | |
| Missing | 319 (0.2) | 37 (0.2) | 282 (0.2) | |

[a]MACU: monitored acute care unit.

[b]SMD: standardized mean difference.

[c]Not applicable.

[d]OB/GYN: obstetrics and gynecology.

[e]ENT: ear, nose, and throat.

## Monitored Acute Care Unit Admissions and Evolution Over Time

Over the study period, 887 (0.5%) patients died in the ED, 97,238 (54.0%) patients were discharged home, 63,025 (35.0%) patients were admitted to a ward, and 19,039 (10.6%) patients were admitted to a monitored acute care unit. Of these, 4499 (23.6%) patients were admitted to the intensive care unit, 10,835 (56.9%) patients were admitted to an intermediate care unit, and 3705 (19.5%) patients were admitted directly to an operating room (Table 2).

**Table .** Patients' disposition from the emergency department by age categories.

| | Total (N=180,189), n (%) | 65 - 74 y (n=61,511), n (%) | 75 - 84 y (n=66,603), n (%) | 85 - 94 y (n=46,890), n (%) | ≥95 y (n=5185), n (%) |
|---|---|---|---|---|---|
| MACU[a] | 19,039 (10.6) | 6890 (11.2) | 7668 (11.5) | 4190 (8.9) | 291 (5.6) |
| Intensive care unit | 4499 (23.6) | 1958 (28.4) | 1917 (25.0) | 611 (14.6) | 13 (4.5) |
| Intermediate care unit | 10,835 (56.9) | 3508 (50.9) | 4339 (56.6) | 2788 (66.5) | 200 (68.7) |
| Operating room | 3705 (19.5) | 1424 (20.7) | 1412 (18.4) | 791 (18.9) | 78 (26.8) |
| Ward | 63,025 (35.0) | 17,917 (29.1) | 23,999 (36.0) | 18,953 (40.4) | 2156 (41.6) |
| Discharged | 97,238 (54.0) | 36,534 (59.4) | 34,608 (52.0) | 23,411 (49.9) | 2685 (51.8) |
| Died in the emergency department | 887 (0.5) | 170 (0.3) | 328 (0.5) | 336 (0.7) | 53 (1.0) |

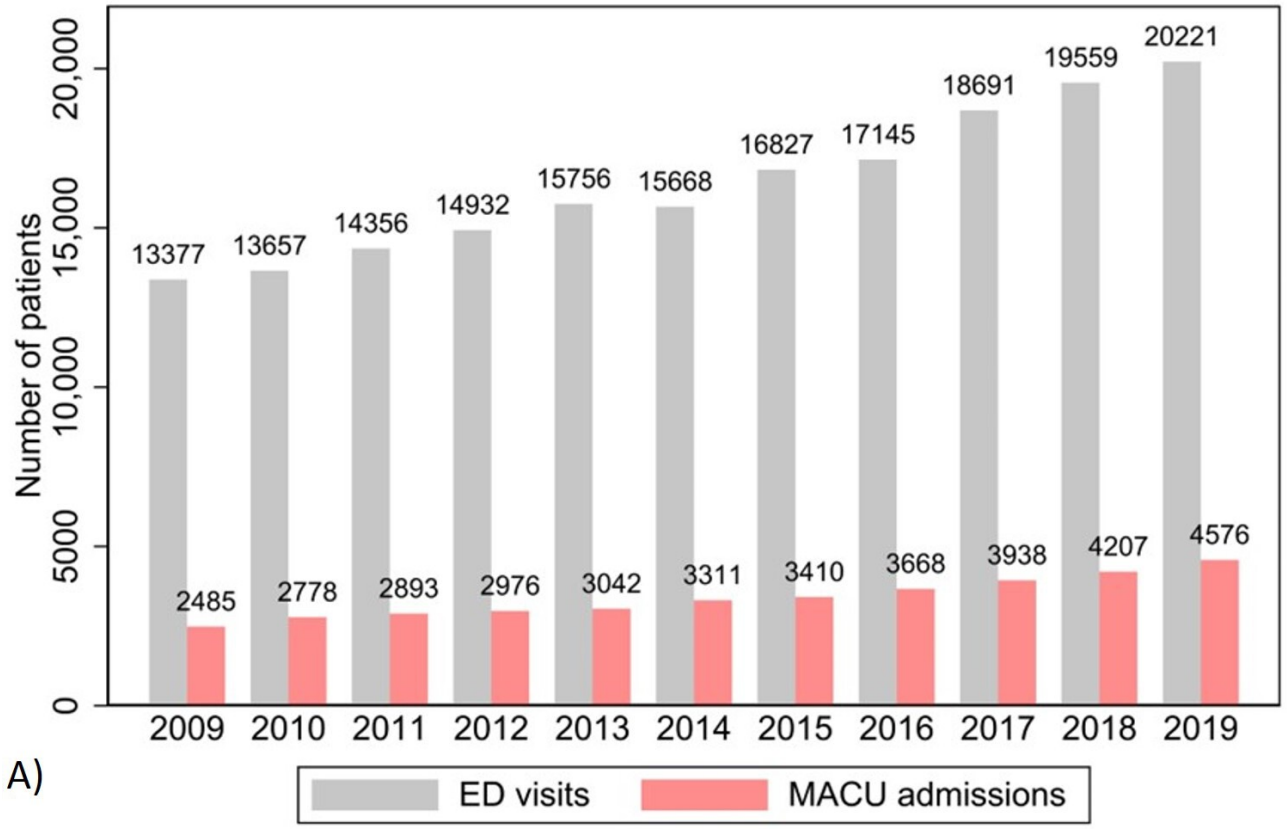
[a]MACU: monitored acute care unit.

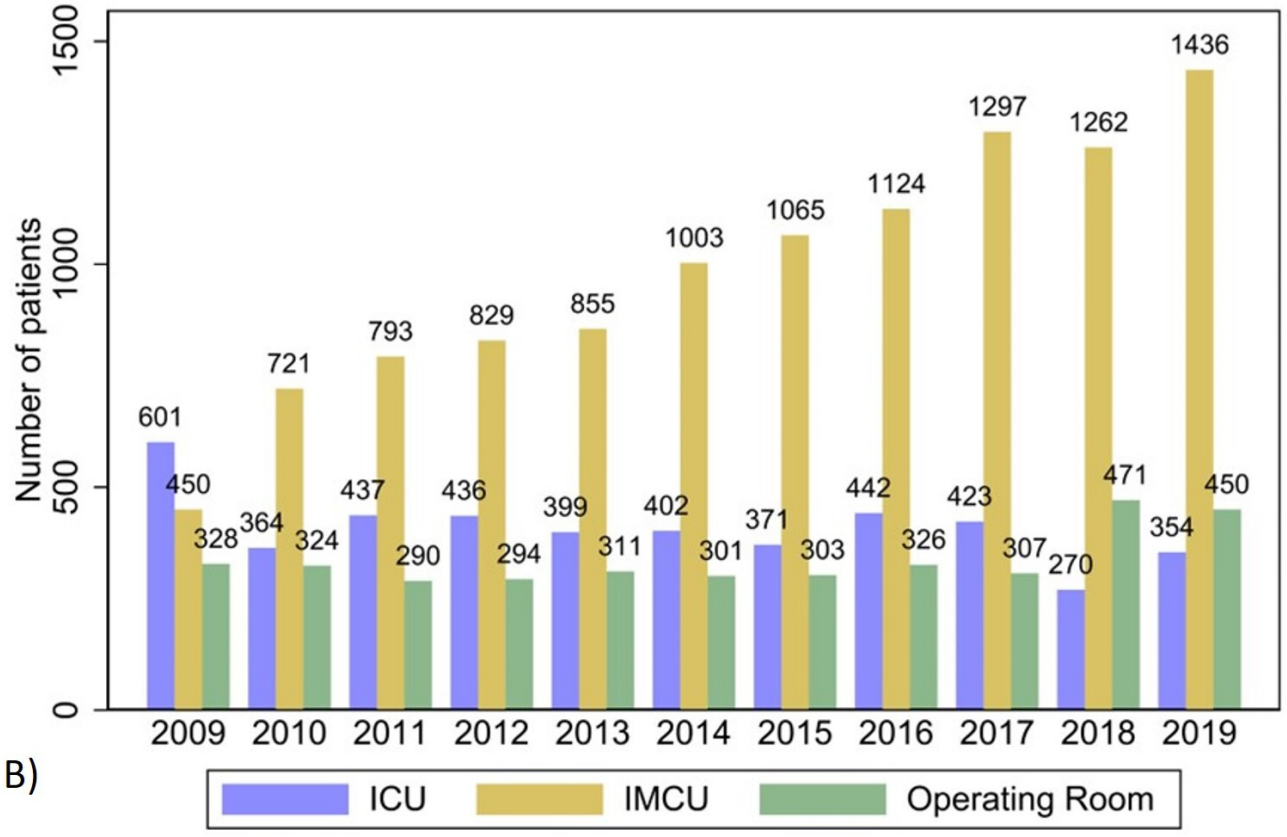
Over the 10-year study period, monitored acute care unit admissions increased by 62.4% in absolute terms (from 1379 in 2009 to 2240 in 2019) and by 0.8% in relative terms (from 10.3% in 2009 to 11.1% in 2019, *P*=.02; Figure 2A). This increase was mainly driven by a rise in intermediate care unit admissions, from 450 of 1379 (32.6%) in 2009 to 1436 of 2240 (64.1%) in 2019. In contrast, operating room admissions remained stable, from 328 of 1379 (23.8%) to 450 of 2240 (20.1%). Intensive care unit admissions decreased, from 601 of 1379 (43.6%) to 354 of 2240 (15.8%; Figure 2B). Notably, while the proportion of patients who died in the ED remained stable, the proportion of patients admitted to hospital wards increased over the 10-year period, from 4026 of 13,377 (30%) in 2009 to 7996 of 19,990 (40%) in 2019.



XSL•FO
RenderX

**Figure 2.** (A) Trends in emergency department (ED) visits and monitored acute care unit (MACU) admissions for patients aged ≥65 years and (B) distribution of MACU admissions to different types of MACUs over time. Over the 10-year study period (2009 - 2019), ED visits at the Geneva University Hospitals from patients aged ≥65 years increased by 56.1%, from 14,705 to 22,955. MACU admissions increased by 62.4%, from 1379 to 2240. The increase in MACU admissions primarily involved intermediate care units (IMCUs), from 450 of 1379 (32.6%) in 2009 to 1436 of 2240 (64.1%) in 2019. ICU: intensive care unit.



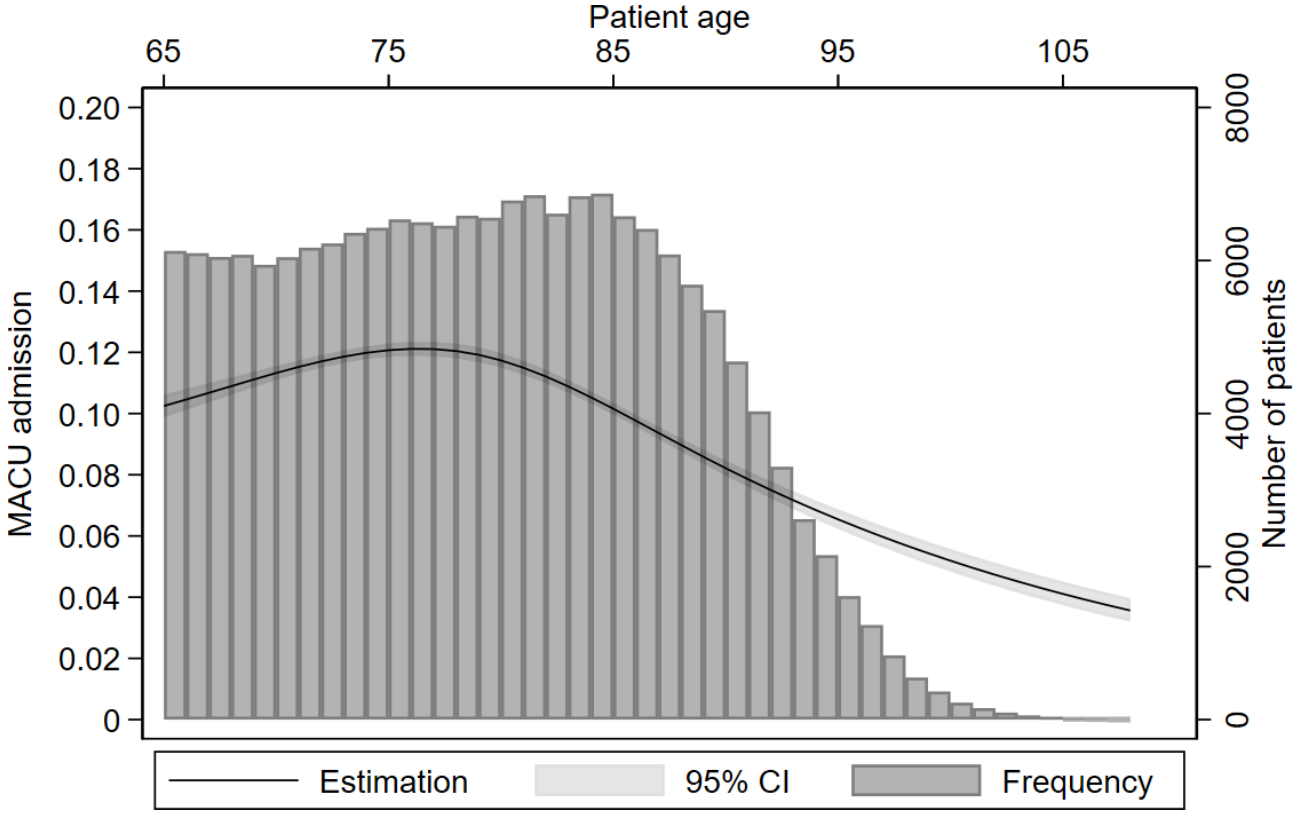



XSL•FO
RenderX

### Predictors for Monitored Acute Care Unit Admission

Patients admitted to monitored acute care units were more frequently male, younger, and more often in a relationship (Table 1). Figure 3 shows the unadjusted association between age and monitored acute care unit admission among older patients. Patients admitted to a monitored acute care unit more frequently arrived by ambulance. They also had a higher triage level, with only 2817 (14.8%) of monitored acute care unit admissions having a lower triage level (3 or 4). In our multivariable analysis, younger age, male sex, marital status, absence of nursing home residence, arrival by ambulance, and higher triage level were all identified as independent predictors of monitored acute care unit admission (Table 3). A sensitivity analysis using age as a categorical variable did not significantly change these findings (Multimedia Appendix 2)

**Figure 3.** Association between age and monitored acute care unit (MACU) admissions among older patients. Older patients are less likely to be admitted to a MACU, with an inflection around 80 to 85 years old. In our multivariable analysis, younger age was identified as an independent predictor of MACU admission (75th vs 25th percentile, adjusted odds ratio 0.73, 95% CI 0.69-0.76).






XSL•FO
RenderX

**Table .** Predictors for monitored acute care unit admission.

| | Multivariable model, adjusted OR[a] (95% CI) |
|---|---|
| Age (y) | |
| 75th versus 25th percentile | 0.73 (0.69-0.76) |
| Sex | |
| Female | Ref.[b] |
| Male | 1.21 (1.14-1.28) |
| Marital status | |
| Married or in a relationship | Ref. |
| Single, divorced, or widowed | 0.88 (0.83-0.93) |
| Nursing home resident | |
| No | Ref. |
| Yes | 0.71 (0.63-0.80) |
| Primary language | |
| French | Ref. |
| Italian | 0.83 (0.74-0.92) |
| Spanish | 0.76 (0.66-0.89) |
| German | 0.97 (0.83-1.14) |
| English | 0.94 (0.78-1.15) |
| Portuguese | 0.81 (0.65-1.02) |
| Others | 0.78 (0.68-0.89) |
| Arrival by ambulance | |
| No | Ref. |
| Yes | 1.87 (1.75-1.99) |
| Triage scale | |
| 1: vital emergency | 39.95 (25.10-63.58) |
| 2: urgent | 9.14 (5.76-14.51) |
| 3: mild urgent | 2.54 (1.60-4.03) |
| 4: nonurgent | Ref. |
| Triage motive category | |
| Cardiology-pneumology | Ref. |
| Neurology-psychiatry | 1.36 (1.28-1.45) |
| Traumatology | 0.61 (0.55-0.67) |
| Digestive-OB/GYN[c] | 1.28 (1.15-1.43) |
| Urology-nephrology | 0.69 (0.56-0.84) |
| Rheumatology | 0.39 (0.28-0.55) |
| Infectious disease | 0.69 (0.58-0.83) |
| Dermatology | 0.57 (0.43-0.75) |
| ENT[d] | 0.35 (0.25-0.49) |
| Others | 1.26 (1.12-1.42) |

[a]OR: odds ratio.

[b]Ref.: reference.

[c]OB/GYN: obstetrics and gynecology.

[d]ENT: ear, nose, and throat.



XSL•FO
RenderX

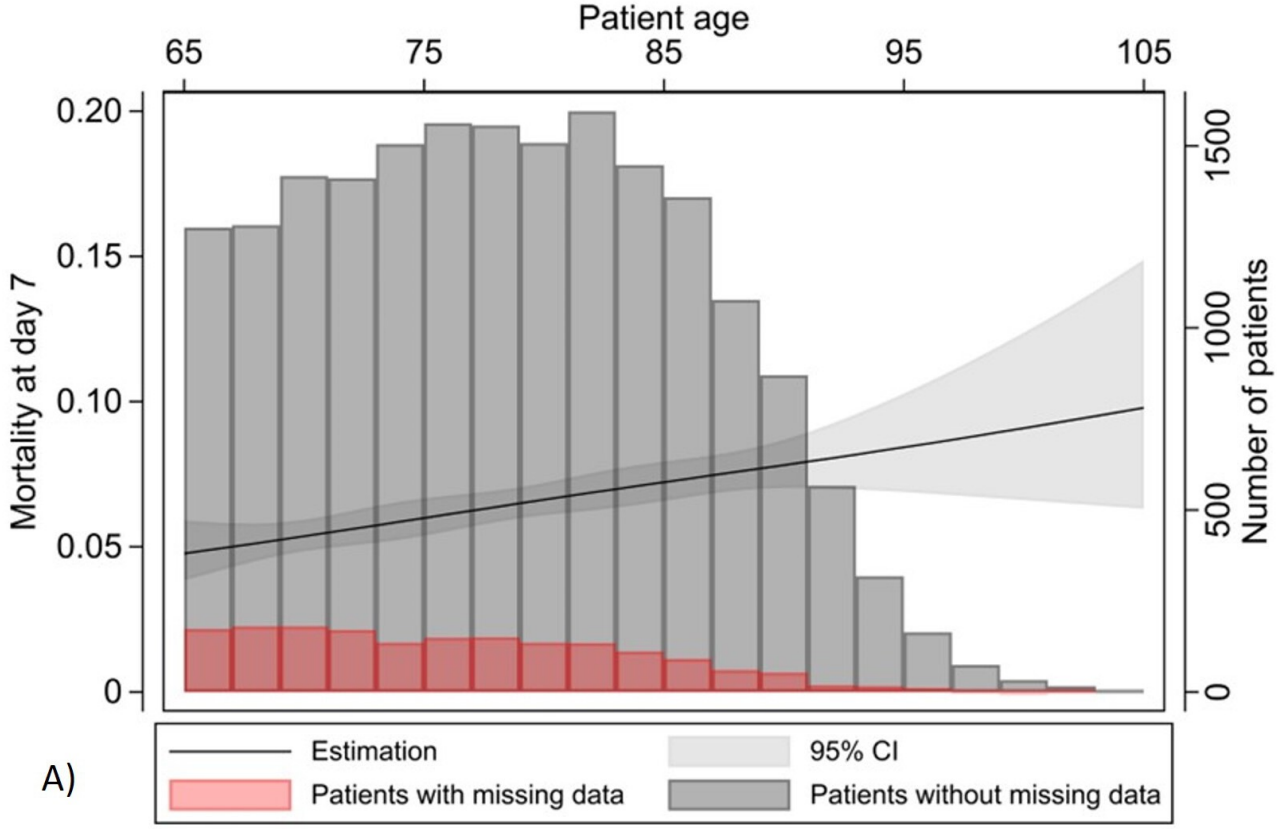
## Outcomes

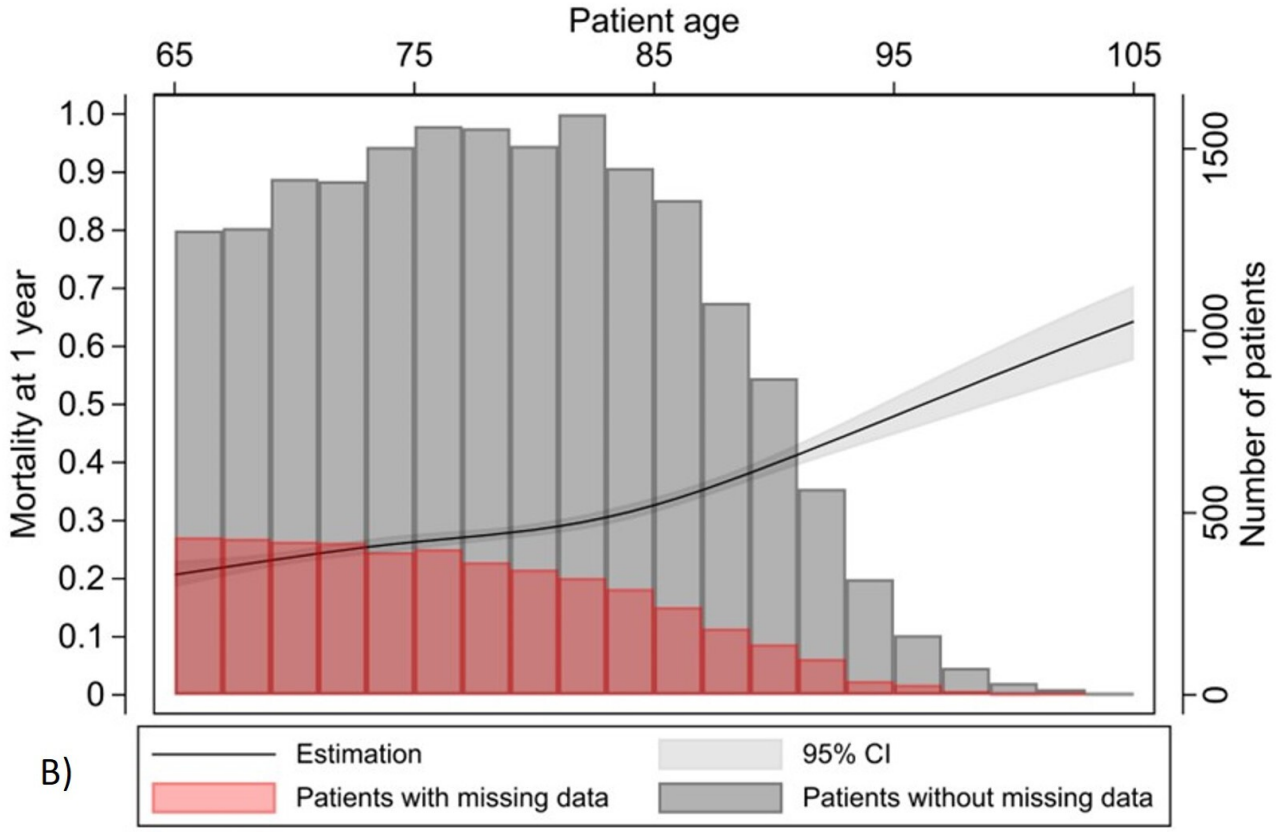
Seven days following monitored acute care unit admission, 1105 (5.8%) patients had died, 6184 (32.5%) patients had been discharged or transferred to long-term care, 10,139 (53.3%) patients were hospitalized in a ward, and 1611 (8.5%) were still in a monitored acute care unit (Table 4). Patients initially admitted to the intensive care unit had the highest mortality (584 patients, 13.0%) and the lowest discharge rate. Only 135 (1.2%) patients admitted to an intermediate care unit required transfer to a higher level of care. The average hospital length of stay ranged from 11.2 to 13.8 days, depending on the admitting unit. One-year mortality following monitored acute care unit admission was 22.3% (4251 patients), with 23.9% (4553 patients) of data missing for the cohort. Compared to patients admitted to an intermediate care unit or the operating room, patients initially admitted to the intensive care unit had worse outcomes (Table 4). After adjustment for other variables, there was a significant association between age and mortality at day 7 (adjusted OR 1.55, 95% CI 1.14-2.10) and at 1 year (adjusted OR 1.28, 95% CI 1.08-1.51), as illustrated in Figure 4.



XSL•FO
RenderX

**Figure 4.** Association between age and (A) mortality at day 7 and (B) mortality at 1 year. For patients aged ≥65 years or older with monitored acute care unit admission, there was a significant association between age and mortality at day 7 (adjusted odds ratio 1.55, 95% CI 1.14-2.10) and at 1 year (adjusted odds ratio 1.28, 95% CI 1.08-1.51). Seven-day mortality was 5.8% (1105 patients), whereas 1-year mortality was 22.3% (4251 patients), with 23.9% (4553 patients) of data missing for the cohort (19,039 patients).






XSL•FO
RenderX

**Table .** Outcomes of patients with monitored acute care unit admission.

| | All (N=19,039) | Intensive care unit (n=4499) | Intermediate care unit (n=10,835) | Operating room (n=3705) |
|---|---|---|---|---|
| Status at day 7, n (%) | | | | |
| Deceased | 1105 (5.8) | 584 (13.0) | 334 (3.1) | 187 (5.0) |
| Intensive care unit | 931 (4.9) | 560 (12.4) | 135 (1.2) | 236 (6.4) |
| Intermediate care unit | 680 (3.6) | 226 (5.0) | 383 (3.5) | 71 (1.9) |
| Ward | 10,139 (53.3) | 2203 (49) | 6173 (57) | 1763 (47.6) |
| Discharged or transferred | 6184 (32.5) | 926 (20.6) | 3810 (35.2) | 1448 (39.1) |
| Hospital length of stay (d), median (IQR) | 9.0 (4.9-15.4) | 10.5 (5.2-18.0) | 8.9 (5.0-14.2) | 8.0 (3.9-15.7) |
| Status at 1 year, n (%) | | | | |
| Alive | 10,230 (53.7) | 2164 (48.1) | 6096 (56.3) | 1970 (53.2) |
| Deceased | 4251 (22.3) | 1355 (30.1) | 2133 (19.7) | 763 (20.6) |
| Unknown | 4553 (23.9) | 980 (21.8) | 2606 (24.1) | 972 (26.2) |

# Discussion

## Principal Findings

This retrospective cohort study, spanning from 2009 to 2019, revealed a significant increase in monitored acute care unit admissions among older patients. Key predictors of monitored acute care unit admission, such as age, sex, and arrival by ambulance, were identified. While short-term outcomes were relatively good with a 7-day mortality rate of approximately 5%, long-term outcomes were poorer, with more than 20% mortality at 1 year.

Over the study period, monitored acute care unit admissions increased by 62%. This increase was not uniform and can be primarily attributed to the growing number of older patients presenting to the ED with a more modest rise in the proportion of these patients being admitted to a monitored acute care unit, from 10.3% (1379 patients) in 2009 to 11.1% (2240 patients) in 2019. Notably, this overall increase in admissions was primarily driven by a substantial rise in intermediate care unit admissions. Although changes in unit capacity over time could not be assessed, the increase in intermediate care unit admissions and the decrease in intensive care unit admissions are likely to be related. This trend may reflect a redistribution of patients and the development of new competencies within the intermediate care units. Such units can also be seen as a more tailored approach to managing older patients, providing an intermediate level of care that is perhaps more suitable for the needs of this population. Additionally, these units benefit from economic advantages, supporting their use as a cost-effective strategy to prioritize health care resource allocation.

Intensive care unit admission rates among older patients have been extensively studied, with the available literature reporting considerable variability. Some studies have found annual increases of up to 5.6% in patients over 80 years old [15], whereas others have found no significant change [16]. In contrast, data on intermediate care unit admissions are more limited, likely due to the wide variety of formats of such units (including intermediate care units, step-down units, and high dependency units), making it challenging to standardize the data [17-19]. The observed rise in intermediate care unit admissions in our study is closely linked to the overall increase in ED visits by older patients, which has directly led to higher demand for monitored care. Beyond this, several other factors may have contributed, including the growing prevalence of chronic diseases among older patients that require a level of care beyond general ward care but less than intensive care unit, along with advances in medical technologies, interventions, protocols, and admission criteria that allow for safer and more effective management of these patients in intermediate care units [20-24].

As anticipated, older patients were less likely to be admitted to a monitored acute care unit, with a notable decrease in admissions and an increase in mortality observed around the age of 85 years. This age threshold may be due to higher comorbidity rates and the presence of advanced care directives in this patient population [25,26]. Age is known to be associated with both a decrease in intensive care unit referral by emergency physicians and an increase in admission denial by intensive care physicians [27]. It might make sense, as intensive care unit admission for patients aged older than 80 years has not been shown to affect 2-year survival rates [28]. However, critical care societies recommend that decisions regarding monitored acute care unit admissions should be based on illness severity, comorbidities, and baseline functional status rather than age alone [29]. It is crucial for physicians to engage in discussions with patients and their next of kin about the potential implications of a monitored acute care unit admission, although the final decision rests with the receiving physician [30].

Our study also identified other predictors of a monitored acute care unit admission. Arrival by ambulance, which is up to 4.6 times more frequent for older patients [31] and serves as an indicator of higher clinical severity, was a strong predictor of monitored acute care unit admission. Nursing home residents were less likely to be admitted to a monitored acute care unit, likely due to higher levels of comorbidities and frailty levels,



XSL•FO
RenderX

often associated with a diminished quality of life and cognitive impairment [32-34]. Neurology-psychiatry–related triage motives were the strongest predictors of monitored acute care unit admission, likely due to conditions such as strokes, which typically require a monitored acute care unit–level care. Interestingly, the French language, male sex, and marital status were also associated with monitored acute care unit admission. Language barriers are well documented as contributing to worse health outcomes, including longer hospital length of stay, higher readmission rates, and limited access to health care systems, regardless of socioeconomic status [35,36]. While female patients were predominant in the ED, they were 21% less likely to be admitted to a monitored acute care unit compared to male patients, a finding consistent with existing literature [37-39]. This disparity may be influenced by more comorbidities in men, and the tendency for female patients to set medical limitations and have advanced directives, especially when divorced or widowed [40]. However, sociocultural factors and implicit or explicit biases may also contribute to this disparity. Unmarried patients often presented with more severe illness at admission [41], possibly due to delayed medical intervention, whereas patients with a partner are more likely to receive high-level care to meet family expectations. Physicians must be aware of such disparities and strive to mitigate them in their practice.

Short-term outcomes for patients admitted to a monitored acute care unit are quite encouraging, with one-third of patients discharged and more than half transferred to the ward by day 7. However, long-term outcomes were less favorable, particularly for patients admitted to the intensive care unit, who had a 1-year mortality rate exceeding 30%. While intensive care unit mortality in older patients is well documented [42], outcomes for intermediate care unit patients have been less studied, especially among ED patients. Torres et al [43] reported no significant difference between demographics regarding in-hospital mortality after intermediate care unit admission and a 34% mortality after 2 years for patients aged older than 65 years, significantly higher than the 10% for younger patients. In a secondary analysis of the ICE-CUB2 trial focusing on ED patients, Thietart et al [44] reported 6-month mortality rates of 44% and 31% for intensive care unit and intermediate care unit patients, respectively, which were higher than those reported in other studies, including this one, which the severity of the patients can explain, as they did not include uncomplicated acute coronary syndromes and strokes. Finally, D'Andrea et al [45] found a 43% 1-year mortality rate in patients aged older than 75 years admitted to a geriatric intermediate care unit, with a significant association between age and 1-year mortality.

## Clinical and Scientific Implications

This study highlights the need to adapt health care structures to meet the growing demands of an aging population. Developing specialized geriatric acute care units or intensive care units, along with training physicians specifically to address the needs of older patients, could help raise awareness of these needs while addressing potential biases related to gender, language, and marital status. Further research should explore patient and family satisfaction and the alignment between care provided and patient wishes. Previous studies have shown that quality of life in older intensive care unit survivors is initially worse [46] but tends to improve within a year, with many older patients willing to undergo intensive care unit admission again if necessary [47,48]. Further studies could focus on developing clinical guidelines or tools specifically tailored to older patients to aid in the shared decision-making process regarding monitored acute care unit admission. These guidelines should incorporate measures such as the Functional Independence Measure or the Clinical Frailty Scale, which have been shown to correlate with resource use and outcomes [45,49,50].

## Strengths and Limitations

Some strengths and limitations need to be acknowledged. The study analyzes a large cohort over a decade, providing a robust dataset for analysis. The innovative and comprehensive approach of considering all monitored acute care unit admissions offers a broad view of high-cost care, which has not been extensively studied. Additionally, we used a rigorous statistical plan, including restricted cubic spline models to account for the nonlinear relationship between age and outcomes, to enhance the robustness of our results. The primary limitations of this study are related to the design. As a retrospective cohort, the study is prone to bias and missing data, notably the variation in monitored acute care unit beds per unit over the years, which may influence patient disposition based on unit saturation. The second major limitation is the lack of information on comorbidities, place of residence (rural vs urban), and polypharmacy. These variables are either not available as structured data in our electronic health record or are at high risk of bias. To preserve high data quality, we chose not to report them. The authors acknowledge this as a limitation of the study, with a risk of residual confounding. Results involving mortality should be considered cautiously, as mortality data were missing for 1 in 4 patients. Our study is monocentric; however, its findings are likely generalizable to other university hospitals in Switzerland and Europe, given the similar monitored acute care unit admission criteria. The exclusion of private hospitals with minimal monitored acute care unit capacity is a minor limitation, as their patients tend to be younger than 65 years. The change in the slope of the association between age and monitored acute care unit admission, as well as age and 1-year mortality, may be driven by only a small number of patients in the oldest age group, potentially distorting the true association. Finally, this study did not cover the periods of the COVID-19 pandemic, which could limit the generalizability of the results, as monitored acute care units were particularly strained during this period.

## Conclusions

This study highlights the sharp increase in monitored acute care unit admissions among older patients, reflecting the growing demand for high levels of care in this population. The rise was primarily driven by a substantial increase in intermediate care unit admissions, whereas intensive care unit admissions remained stable. These findings emphasize the need for hospitals to adapt health care infrastructure, clinical procedures, and resource allocation to meet the evolving needs of an aging population. Future work should focus on the importance of developing specialized geriatric acute care units, refining clinical guidelines, and engaging in shared decision-making to ensure equitable and effective care for all older patients.



XSL•FO
RenderX

## Acknowledgments

Language and syntax were reviewed and corrected using artificial intelligence–based tools to improve clarity and consistency of the manuscript. The authors declare the use of generative artificial intelligence (GAI) in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GAI tools under full human supervision: proofreading and editing, adapting and adjusting emotional tone, and reformatting. The GAI tool used was ChatGPT. Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes.

## Funding

There was no specific funding for this study.

## Data Availability

The datasets generated or analyzed during this study are available with this manuscript (Multimedia Appendix 3).

## Authors' Contributions

Conceptualization: LvD, OG, CAF. Data curation: CAF. Formal analysis: CAF. Funding acquisition: TD. Investigation: CAF. Methodology: LvD, SvD, CAF. Project administration: CAF. Supervision: CAF. Visualization: LvD, CAF. Writing – original draft : LvD, SvD, CAF. Writing – review & editing : LvD, SvD, FR, AR, CM, SC, XR, TD, OG, CAF.

## Conflicts of Interest

SvD received speaking honoraria from Löwenstein Medical. All other authors declare no competing interests.

Multimedia Appendix 1
Proportion of ED visits by older patients. Between 2009 and 2019, a total of 701,838 ED visits were recorded, with a global increase of 34.1% (56,944-76,368) in all ED visits. This increase was more pronounced for patients aged ≥65 years (14,705-22,955, +56.1%) than for patients aged <65 years (42,239 to 53,413, +26.5%; *P*<.001). ED: emergency department.
[PNG File, 40 KB - i-jmr_v15i1e80629_app1.png ]

Multimedia Appendix 2
Potential predictors for MACU admission with age in categories. MACU: monitored acute care unit.
[DOCX File, 108 KB - i-jmr_v15i1e80629_app2.docx ]

Multimedia Appendix 3
Dataset for publication. To ensure patient confidentiality, the dataset prepared for publication excludes the patient's first language, exact admission date, and identifying information, such as patient ID.
[ZIP File, 2865 KB - i-jmr_v15i1e80629_app3.zip ]

## References

1. Ukkonen M, Jämsen E, Zeitlin R, Pauniaho SL. Emergency department visits in older patients: a population-based survey. BMC Emerg Med 2019 Feb 27;19(1):20. [doi: 10.1186/s12873-019-0236-3] [Medline: 30813898]
2. Wittenberg R, Sharpin L, McCormick B, Hurst J. The ageing society and emergency hospital admissions. Health Policy 2017 Aug;121(8):923-928. [doi: 10.1016/j.healthpol.2017.05.007] [Medline: 28619464]
3. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012 Jul 7;380(9836):37-43. [doi: 10.1016/S0140-6736(12)60240-2] [Medline: 22579043]
4. Albert M, McCaig LF, Ashman JJ. Emergency department visits by persons aged 65 and over: United States, 2009-2010. : Centers for Disease Control and Prevention, National Center for Health Statistics; 2013 Oct URL: https://www.cdc.gov/nchs/data/databriefs/db130.pdf [accessed 2026-01-17]
5. Liston P, Conyngham G, Brady M, Byrne PJ, Gilligan P. Growing old in the emergency department. Ir Med J 2017 Aug 12;110(7):621. [Medline: 29169003]
6. Latham LP, Ackroyd-Stolarz S. Emergency department utilization by older adults: a descriptive study. Can Geriatr J 2014 Dec;17(4):118-125. [doi: 10.5770/cgj.17.108] [Medline: 25452824]
7. Wunsch H, Guerra C, Barnato AE, Angus DC, Li G, Linde-Zwirble WT. Three-year outcomes for Medicare beneficiaries who survive intensive care. JAMA 2010 Mar 3;303(9):849-856. [doi: 10.1001/jama.2010.216] [Medline: 20197531]



XSL•FO
RenderX

8. Fuchs L, Chronaki CE, Park S, et al. ICU admission characteristics and mortality rates among elderly and very elderly patients. Intensive Care Med 2012 Oct;38(10):1654-1661. [doi: 10.1007/s00134-012-2629-6] [Medline: 22797350]
9. Somme D, Maillet JM, Gisselbrecht M, Novara A, Ract C, Fagon JY. Critically ill old and the oldest-old patients in intensive care: short- and long-term outcomes. Intensive Care Med 2003 Dec;29(12):2137-2143. [doi: 10.1007/s00134-003-1929-2] [Medline: 14614546]
10. Dowdy DW, Eid MP, Sedrakyan A, et al. Quality of life in adult survivors of critical illness: a systematic review of the literature. Intensive Care Med 2005 May;31(5):611-620. [doi: 10.1007/s00134-005-2592-6] [Medline: 15803303]
11. Rutschmann OT, Hugli OW, Marti C, et al. Reliability of the revised Swiss emergency triage scale: a computer simulation study. Eur J Emerg Med 2018 Aug;25(4):264-269. [doi: 10.1097/MEJ.0000000000000449] [Medline: 28099182]
12. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996 Dec;49(12):1373-1379. [doi: 10.1016/s0895-4356(96)00236-3] [Medline: 8970487]
13. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol 2007 Mar 15;165(6):710-718. [doi: 10.1093/aje/kwk052] [Medline: 17182981]
14. Harrell FE. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis: Springer International Publishing; 2015. [doi: 10.1007/978-3-319-19425-7]
15. Bagshaw SM, Webb SAR, Delaney A, et al. Very old patients admitted to intensive care in Australia and New Zealand: a multi-centre cohort analysis. Crit Care 2009;13(2):R45. [doi: 10.1186/cc7768] [Medline: 19335921]
16. Rai S, Brace C, Ross P, et al. Characteristics and outcomes of very elderly patients admitted to intensive care: a retrospective multicenter cohort analysis. Crit Care Med 2023 Oct 1;51(10):1328-1338. [doi: 10.1097/CCM.0000000000005943] [Medline: 37219961]
17. Plate JDJ, Leenen LPH, Houwert M, Hietbrink F. Utilisation of intermediate care units: a systematic review. Crit Care Res Pract 2017;2017:8038460. [doi: 10.1155/2017/8038460] [Medline: 28775898]
18. López-Jardón P, Martínez-Fernández MC, García-Fernández R, Martín-Vázquez C, Verdeal-Dacal R. Utility of intermediate care units: a systematic review study. Healthcare (Basel) 2024 Jan 24;12(3):296. [doi: 10.3390/healthcare12030296] [Medline: 38338181]
19. Prin M, Wunsch H. The role of stepdown beds in hospital care. Am J Respir Crit Care Med 2014 Dec 1;190(11):1210-1216. [doi: 10.1164/rccm.201406-1117PP] [Medline: 25163008]
20. Chou YS, Lin HY, Weng YM, et al. Step-down units are cost-effective alternatives to coronary care units with non-inferior outcomes in the management of ST-elevation myocardial infarction patients after successful primary percutaneous coronary intervention. Intern Emerg Med 2020 Jan;15(1):59-66. [doi: 10.1007/s11739-019-02037-z] [Medline: 30706252]
21. Meaudre E, Nguyen C, Contargyris C, et al. Management of septic shock in intermediate care unit. Anaesth Crit Care Pain Med 2018 Apr;37(2):121-127. [doi: 10.1016/j.accpm.2017.07.004] [Medline: 28790011]
22. Plate JDJ, Peelen LM, Leenen LPH, Hietbrink F. The intermediate care unit as a cost-reducing critical care facility in tertiary referral hospitals: a single-centre observational study. BMJ Open 2019 Jun 4;9(6):e026359. [doi: 10.1136/bmjopen-2018-026359] [Medline: 31167865]
23. Ranhoff AH, Rozzini R, Sabatini T, et al. Subintensive care unit for the elderly: a new model of care for critically ill frail elderly medical patients. Intern Emerg Med 2006;1(3):197-203. [doi: 10.1007/BF02934737] [Medline: 17120465]
24. Ohbe H, Matsui H, Yasunaga H. Intensive care unit versus high-dependency care unit for patients with acute heart failure: a nationwide propensity score-matched cohort study. J Intensive Care 2021 Dec 20;9(1):78. [doi: 10.1186/s40560-021-00592-2] [Medline: 34930470]
25. Gamertsfelder EM, Seaman JB, Tate J, Buddadhumaruk P, Happ MB. Prevalence of advance directives among older adults admitted to intensive care units and requiring mechanical ventilation. J Gerontol Nurs 2016 Apr;42(4):34-41. [doi: 10.3928/00989134-20151124-02] [Medline: 26651862]
26. Foley C, Bloomer M, Hutchinson AM. Factors that influence intensive care admission decisions for older people: a systematic review. Aust Crit Care 2023 Mar;36(2):274-284. [doi: 10.1016/j.aucc.2021.12.006] [Medline: 35144889]
27. Guidet B, Boumendil A, Garrousteorgeas M, Pateron D. Admission en réanimation du sujet âgé à partir du service des urgences. État des lieux [Article in French]. Réanimation 2008 Dec;17(8):790-801. [doi: 10.1016/j.reaurg.2008.09.004]
28. Abuhasira R, Anstey M, Novack V, Bose S, Talmor D, Fuchs L. Intensive care unit capacity and mortality in older adults: a three nations retrospective observational cohort study. Ann Intensive Care 2022 Mar 4;12(1):20. [doi: 10.1186/s13613-022-00994-x] [Medline: 35244803]
29. Nates JL, Nunnally M, Kleinpell R, et al. ICU admission, discharge, and triage guidelines: a framework to enhance clinical operations, development of institutional policies, and further research. Crit Care Med 2016 Aug;44(8):1553-1602. [doi: 10.1097/CCM.0000000000001856] [Medline: 27428118]
30. Blanch L, Abillama FF, Amin P, et al. Triage decisions for ICU admission: report from the Task Force of the World Federation of Societies of Intensive and Critical Care Medicine. J Crit Care 2016 Dec;36:301-305. [doi: 10.1016/j.jcrc.2016.06.014] [Medline: 27387663]
31. Baum SA, Rubenstein LZ. Old people in the emergency room: age-related differences in emergency department use and care. J Am Geriatr Soc 1987 May;35(5):398-404. [doi: 10.1111/j.1532-5415.1987.tb04660.x] [Medline: 3571788]




XSL•FO
RenderX

32. Manis DR, Katz P, Lane NE, et al. Clinical comorbidities and transitions between care settings among residents of assisted living facilities: a repeated cross-sectional study. J Am Med Dir Assoc 2023 Sep;24(9):1356-1360. [doi: 10.1016/j.jamda.2023.06.017] [Medline: 37507099]
33. Moore KL, Boscardin WJ, Steinman MA, Schwartz JB. Patterns of chronic co-morbid medical conditions in older residents of U.S. nursing homes: differences between the sexes and across the agespan. J Nutr Health Aging 2014 Apr;18(4):429-436. [doi: 10.1007/s12603-014-0001-y] [Medline: 24676326]
34. Toot S, Swinson T, Devine M, Challis D, Orrell M. Causes of nursing home placement for older people with dementia: a systematic review and meta-analysis. Int Psychogeriatr 2017 Feb;29(2):195-208. [doi: 10.1017/S1041610216001654] [Medline: 27806743]
35. Karliner LS, Kim SE, Meltzer DO, Auerbach AD. Influence of language barriers on outcomes of hospital care for general medicine inpatients. J Hosp Med 2010;5(5):276-282. [doi: 10.1002/jhm.658] [Medline: 20533573]
36. Lindholm M, Hargraves JL, Ferguson WJ, Reed G. Professional language interpretation and inpatient length of stay and readmission rates. J Gen Intern Med 2012 Oct;27(10):1294-1299. [doi: 10.1007/s11606-012-2041-5] [Medline: 22528618]
37. Lat TI, McGraw MK, White HD. Gender differences in critical illness and critical care research. Clin Chest Med 2021 Sep;42(3):543-555. [doi: 10.1016/j.ccm.2021.04.012] [Medline: 34353458]
38. Merdji H, Long MT, Ostermann M, et al. Sex and gender differences in intensive care medicine. Intensive Care Med 2023 Oct;49(10):1155-1167. [doi: 10.1007/s00134-023-07194-6] [Medline: 37676504]
39. Modra L, Pilcher D, Bailey M, Bellomo R. Sex differences in intensive care unit admissions in Australia and New Zealand. Crit Care Resusc 2021 Mar;23(1):86-93. [doi: 10.51893/2021.1.OA8] [Medline: 38075190]
40. McPherson K, Carlos WG, Emmett TW, Slaven JE, Torke AM. Limitation of life-sustaining care in the critically ill: a systematic review of the literature. J Hosp Med 2019 May;14(5):303-310. [doi: 10.12788/jhm.3137] [Medline: 30794145]
41. Gordon HS, Rosenthal GE. Impact of marital status on outcomes in hospitalized patients. Evidence from an academic medical center. Arch Intern Med 1995;155(22):2465-2471. [Medline: 7503606]
42. Vallet H, Schwarz GL, Flaatten H, de Lange DW, Guidet B, Dechartres A. Mortality of older patients admitted to an ICU: a systematic review. Crit Care Med 2021 Feb 1;49(2):324-334. [doi: 10.1097/CCM.0000000000004772] [Medline: 33332816]
43. Torres OH, Francia E, Longobardi V, Gich I, Benito S, Ruiz D. Short- and long-term outcomes of older patients in intermediate care units. Intensive Care Med 2006 Jul;32(7):1052-1059. [doi: 10.1007/s00134-006-0170-1] [Medline: 16791668]
44. Thietart S, Boumendil A, Pateron D, Guidet B, Vallet H, ICE-CUB2 Study Network. Impact on 6-month outcomes of hospital trajectory in critically ill older patients: analysis of the ICE-CUB2 clinical trial. Ann Intensive Care 2022 Jul 11;12(1):65. [doi: 10.1186/s13613-022-01042-4] [Medline: 35819563]
45. D'Andrea A, Le Peillet D, Fassier T, et al. Functional Independence Measure score is associated with mortality in critically ill elderly patients admitted to an intermediate care unit. BMC Geriatr 2020 Sep 9;20(1):334. [doi: 10.1186/s12877-020-01729-y] [Medline: 32907534]
46. Ariyo K, Canestrini S, David AS, Ruck Keene A, Wolfrum S, Owen G. Quality of life in elderly ICU survivors before the COVID-19 pandemic: a systematic review and meta-analysis of cohort studies. BMJ Open 2021 Oct 11;11(10):e045086. [doi: 10.1136/bmjopen-2020-045086] [Medline: 34635510]
47. Tabah A, Philippart F, Timsit JF, et al. Quality of life in patients aged 80 or over after ICU discharge. Crit Care 2010;14(1):R2. [doi: 10.1186/cc8231] [Medline: 20064197]
48. Israelsson-Skogsberg Å, Eriksson T, Lindberg E. A scoping review of older patients' health-related quality of life, recovery and well-being after intensive care. Nurs Open 2023 Sep;10(9):5900-5919. [doi: 10.1002/nop2.1873] [Medline: 37306357]
49. Fernando SM, McIsaac DI, Perry JJ, et al. Frailty and associated outcomes and resource utilization among older ICU patients with suspected infection. Crit Care Med 2019 Aug;47(8):e669-e676. [doi: 10.1097/CCM.0000000000003831] [Medline: 31135504]
50. Fehlmann CA, Nickel CH, Cino E, Al-Najjar Z, Langlois N, Eagles D. Frailty assessment in emergency medicine using the Clinical Frailty Scale: a scoping review. Intern Emerg Med 2022 Nov;17(8):2407-2418. [doi: 10.1007/s11739-022-03042-5] [Medline: 35864373]


## Abbreviations

**ED:** emergency department
**HUG:** Hôpitaux Universitaires de Genève/Geneva University Hospitals
**OR:** odds ratio
**SMD:** standardized mean difference



XSL•FO
RenderX

*Edited by M Balcarras, N Cahill; submitted 22.Jul.2025; peer-reviewed by C Thompson; revised version received 27.Dec.2025; accepted 29.Dec.2025; published 02.Feb.2026.*

*Please cite as:*
*van Delden L, von Düring S, Rouyer F, Rossel A, Marti C, Cereghetti S, Roux X, Desmettre T, Grosgurin O, Fehlmann CA*
*Trends, Predictors, and Outcomes of Monitored Acute Care Unit Admissions in Older Adults: 10-Year Retrospective Analysis*
*Interact J Med Res 2026;15:e80629*
*URL: https://www.i-jmr.org/2026/1/e80629*
*doi:10.2196/80629*

© Ludovic van Delden, Stephan von Düring, Frederic Rouyer, Anne Rossel, Christophe Marti, Sara Cereghetti, Xavier Roux, Thibaut Desmettre, Olivier Grosgurin, Christophe A Fehlmann. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 2.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Frailty Screening and Management for Older Australians in General Practice: Mixed Methods Evaluation

Jennifer R Job[1], PhD; Caroline Nicholson[1], PhD; Ruby Strauss[1], MPhil; Debra Clark[2]; Anita Pelecanos[3], BHSc(Hon); Claire Jackson[1,4], MD

[1]Centre for Health System Reform and Integration, Mater Research-The University of Queensland, Level 8, Health Sciences Building, RBWH, Brisbane, Australia

[2]Aged Care, Sydney North Health Network, Sydney, Australia

[3]Statistics Unit, QIMR Berghofer Medical Research Institute, Brisbane, Australia

[4]General Practice Clinical Unit, The University of Queensland, Brisbane, Australia

**Corresponding Author:**
Jennifer R Job, PhD
Centre for Health System Reform and Integration, Mater Research-The University of Queensland, Level 8, Health Sciences Building, RBWH, Brisbane, Australia

## Abstract

**Background:** Frailty increases with age and is associated with increased vulnerability to adverse health outcomes. International guidelines recommend screening for frailty in primary care; however, this is not routine practice in Australia. Once identified, frailty progression has the potential to be halted or reversed with early intervention. The FRAIL (Fatigue, Resistance, Ambulation, Illnesses, Loss of weight) Scale Tool, a simple and validated screening and management tool, offers a feasible approach for integration into the Australian health assessment for those aged 75 years and older (75+HA), which can be performed annually by primary care providers.

**Objective:** This study explores the rates of frailty, resources required to support management, and the determinants of implementing frailty screening and providing management for older Australians at the 75+HA.

**Methods:** A mixed methods evaluation was conducted in 24 general practices across 2 Australian Primary Health Network regions, Sydney North and Brisbane South. The FRAIL Scale Tool was implemented during the 75+ health assessment, and data were collected on FRAIL Scale scores, hospitalization rates, recommended frailty interventions, and barriers to frailty management. Practice staff perceptions of the long-term sustainment of the FRAIL Scale Tool were assessed using the Provider Report of Sustainment Scale. Semistructured qualitative interviews were conducted with practice staff and patients, exploring barriers and enablers to implementing frailty screening and management. Guided by the Consolidated Framework for Implementation Research, transcripts were coded and themes developed.

**Results:** Of the 1484 patients aged ≥75 years who were screened, 223 (15%) patients were frail, 616 (41.5%) patients were prefrail, and 645 (43.5%) patients were robust. People who were frail were more likely to be female, older, and have more prescribed medications. Of those screened as frail, 23 (11%) had a nonelective hospitalization in the 3 months prior to screening compared with 28 (5%) who screened as prefrail and 5 (1%) who screened as robust ($P$=.012). Management recommendations commonly included medication reviews, aged care packages, assessment for depression, and exercise programs. Barriers identified to accessing interventions included health, transport, cost, and time. Survey and qualitative findings highlighted that the FRAIL Scale Tool was easy to use, integrated well into existing workflows as part of the 75+HA, and sustained use would be supported by software integration. Patients valued the assessment and tailored health support offered by trusted primary care providers.

**Conclusions:** Incorporating the FRAIL Scale Tool into the annual health assessment for people aged 75 years and older provides a funded opportunity for addressing frailty in general practice. Patients and staff value the Tool's simplicity and the opportunity to raise awareness and manage frailty proactively. Incorporating the Tool into practice software systems would enhance adoption. Broader implementation research in diverse settings and with Aboriginal and Torres Strait Islander populations is needed to improve frailty prevention and management.

***(Interact J Med Res 2026;15:e79681)*** doi:10.2196/79681

**KEYWORDS**

FRAIL scale; implementation; older adults; frailty screening; primary care



XSL•FO
RenderX

## Introduction

Worldwide, the population is aging with an increasing proportion of people aged 60 years and older, and the number of people aged 80 years or older is expected to triple between 2020 and 2050 [1]. In response, health care systems and models of care must focus on the health needs and priorities of older people. Frailty and its precursor, prefrailty, affect more than half of the adults older than 50 years worldwide [2]. Characterized by physiological decline and increased vulnerability to stressors, frailty is a dynamic state associated with an increased risk of adverse health outcomes [3]. While not an inevitable part of aging, the risk of developing frailty increases with advancing age, with the prevalence of frailty and prefrailty climbing to 31% and 52% respectively, in people older than 80 years [2]. With the gradual loss of muscle mass and decline in physical condition, frailty is associated with falls and acute illnesses. Hospitalization or prolonged recovery can contribute to further physical deconditioning, with people experiencing frailty often requiring substantial support and resources to return to their prior health state or prevent further deterioration [4]. Importantly, causes of frailty can be managed, and in some cases reversed, highlighting the importance of early identification of older people who are living with prefrailty [5].

International guidelines highlight the opportunity for frailty prevention and early intervention in primary care [6]. This can be challenging as prefrailty develops subtly and may go unnoticed without targeted examination. While screening of older adults in primary care has been recommended for early identification of prefrailty to initiate preventative management strategies [7], screening is not routinely recorded and tracked in practice in Australian primary care [8]. This presents a missed opportunity as validated screening tools have been found to assist health care professionals with identifying risk of frailty and monitoring any changes in health status [6].

The FRAIL (Fatigue, Resistance, Ambulation, Illnesses, Loss of weight) Scale (Figure 1) has been validated [9,10] as a screening tool for community-dwelling adults. A recent pilot study found the Tool was feasible for general practitioners (GPs) and practice nurses to administer in a busy clinical setting and appropriate for screening older people at the time of routine appointments [11]. It is based on the phenotype of frailty elements of Fried et al [12]: exhaustion (fatigue), weight loss, measured grip strength and walking speed, and low energy expenditure. The FRAIL Scale includes 5 simple questions about fatigue, resistance, ambulation, illnesses, and loss of weight to identify risk of frailty [9]. FRAIL Scale scores range from 0 to 5 (with 1 point for each component) to give one of three outcomes: (1) frail (3 or more components present), (2) prefrail (1 to 2 components present), or (3) robust (0 components present).



XSL•FO
RenderX

**Figure 1.** The FRAIL Scale. FRAIL: Fatigue, Resistance, Ambulation, Illnesses, Loss of weight.

**FRAIL SCALE Risk Assessment**

| | Question | Scoring | Result |
|---|---|---|---|
| F | **Fatigue**<br>How much of the time during the past 4 weeks did you feel tired?<br>A = All or most of the time<br>B = Some, a little or none of the time | A = 1<br>B = 0 | |
| R | **Resistance**<br>In the last 4 weeks b y yourself and not using aids, do you have any difficulty walking up 10 steps without resting? | Yes = 1<br>No = 0 | |
| A | **Ambulation**<br>In the last 4 weeks by yourself and not using aids , do you have any difficulty walking 300 meters? | Yes = 1<br>No = 0 | |
| I | **Illness**<br>Did your Doctor ever tell you that you have?<br>☐ Hypertension<br>☐ Diabetes<br>☐ Cancer (not a minor skin cancer)<br>☐ Chronic lung disease<br>☐ Heart Attack<br>☐ Congestive heart failure<br>☐ Angina<br>☐ Asthma<br>☐ Arthritis<br>☐ Stroke<br>☐ Kidney disease | 0 – 4 answers ✓ = 0<br>5 – 11 answers ✓ = 1 | |
| L | **Loss of weight**<br>Have you unintentionally lost more than 5kg or 5% of your body weight in the past year? | Yes = 1<br>No = 0 | |
| | | **Total Score** | |
| **Scoring: Robust = 0, Pre-frail = 1-2, Frail = ≥3** | | | |

Once identified, frailty progression has the potential to be halted or reversed with proper management [5]. Clinical guidelines suggest a multifaceted management approach to address physical function, nutrition, polypharmacy, and fatigue [6,13,14]. However, according to health professionals, the management of frailty can be hindered by a lack of defined protocols and care pathways to guide evidence-based intervention in practice [15,16]. To address this issue, the bespoke Frailty Management/Decision Tool (the Tool) was developed by Sydney North Health Network (SNHN) and Northern Sydney Local Health District for primary care staff to assess and track FRAIL Scale scores and provide associated management suggestions and interventions recommended within the Tool (Figure 2). The tool was successfully piloted by general practices in Sydney North and Brisbane South, with 80% (860/1071) of those screened identified as frail or prefrail [11]. Staff feedback



XSL•FO
RenderX

highlighted that the Tool was patient-focused, easy to use, and a formalized assessment for risk of frailty. To fit with practice workflows, this pilot research supported screening for frailty risk in conjunction with the annual Medicare-funded preventive health assessment for those aged 75 years and older (75+HA), which is conducted with more than 600,000 older Australians in general practice annually [17]. While the 75+HA assesses falls risk, vision, hearing, current weight, and cognitive assessment, it does not currently include a formalized frailty assessment.



XSL•FO
RenderX

**Figure 2.** The Frailty Management Decision Tool (The FRAIL Scale Tool).

## FRAILTY MANAGEMENT/DECISION SUPPORT TOOL

| | Assessment Score | Intervention | Referral/follow up |
|---|---|---|---|
| | FRAIL scale<br>0= robust | • Encourage ongoing activity levels.<br>• Provide physical activity and nutrition resources | • Re-do FRAIL Scale in 12 months<br>• Community exercise with balance/resistance component. Provide physical activity and nutrition resources. |
| | FRAIL scale<br>1-2 = Pre-frail<br>>3 = Frail | **If FRAIL Score is positive, address underlying causes as suggested below** | |
| F | Feeling fatigued most or all of the time | • Consider screening for reversible causes of fatigue (sleep apnoea, depression, anaemia, hypotension, hypothyroidism, B12 deficiency).<br>• Use EPWORTH scale, K10 or Geriatric Depression Scale in Health Assessment. | • Consider referral to Geriatrician/Specialist for complex care patients<br>• Consider referral to Occupational Therapist for functional and home review<br>• Consider referral to Psychologist using Mental Health Care Plan<br>• Consider referral to Aged Care organisation for loneliness support (isolation can be a cause of fatigue). |
| R | Resistance against gravity - difficulty walking up 10 steps without resting | • Consider referring to an individualized progressive exercise program with resistance and strength component. | • Physiotherapist or exercise physiologist for exercise prescription.<br>• For diabetes – group session Medicare funded exercise physiologist<br>• Community exercise with balance/resistance component.<br>• Telephone-based health coaching |
| A | Ambulation - difficulty walking 300 meters unaided | | |
| I | Having 5 or more illnesses | • Review indication, side effects and use of medication (evidence for use of some medicines change after 75).<br>• Consider discussing with pharmacist.<br>• Consider reducing/de-prescribing superfluous medication. | • Pharmacist for comprehensive medication review (HMR item 900).<br>• Occupational Therapist for functional and home safety review.<br>• Self-management support from aged care organisation volunteer |
| L | Loss of >5% weight in 12 months | • Consider screening for reversible causes of weight loss<br>• Consider protein and caloric supplementation/food fortification (75mg protein per day required – range of products available at pharmacy).<br>• Advice and encourage healthy eating; provide nutrition resources. | • Weight and assess BMI – record in patient record<br>• Dietitian for diet review and management<br>• Meal Delivery Services<br>• Speech Pathologist for swallowing review<br>• Dentist for dental review (pain/infection/ill-fitting dentures)<br>• Occupational Therapist for functional and home cooking ability/review. |

This study aims to determine:



XSL•FO
RenderX

1. The rates of risk of frailty/frailty in the general practice population aged 75 years and older when screened at the time of the 75+HA.
2. The resources required to support those identified as frail or at risk of frailty.
3. Barriers and enablers to implementing frailty screening and management for community-dwelling older Australians in primary care.

## Methods

### Design

A convergent mixed methods study design was used. FRAIL Scale scores, recommended resources for frailty management, and barriers to accessing these were tracked prospectively by general practice staff. Semistructured qualitative interviews were conducted with practice staff and patients to understand (1) the determinants of implementing frailty screening and (2) the barriers and enablers to frailty management in primary care. Interview questions and analysis were guided by the Consolidated Framework for Implementation Research (CFIR) [18]. CFIR provides a framework for assessing determinants of implementation at multiple contextual domain levels: inner and outer setting, intervention, individual, and implementation process.

### Researcher Positionality

DC is a Primary Health Network (PHN) stakeholder who worked with Sydney North primary care centers to introduce the FRAIL Scale tool. AP contributed as a biostatistician, and the remainder of the research team has worked clinically with older persons (JJ: Dietitian, CN: Physiotherapist, RS: Sleep Technician, CJ: GP). All have extensive experience implementing health services in the primary care setting and are acutely aware of the time and workforce challenges faced by primary care.

### Ethical Considerations

Ethics was approved by The University of Queensland HREC (Project IDs: 2021/HE001441 and 2021/HE002398), and informed consent was obtained from participants. Participant data was protected by unique study codes and secure servers with limited access to ensure confidentiality, and only deidentified data was shared with the researchers. General practice incentive payments supported the general practice staff time for the data collection (70 frailty screens/practice at Aus $40 [US $26.652]/assessment for new practices). Participants were reimbursed for their time for a semistructured interview at Aus $120 (US $79.956)/interview for practice staff and Aus $50 (US $33.315) for patients.

### Recruitment

As previously described, the research team partnered with 2 PHNs—SNHN and Brisbane South Primary Health Network (BSPHN)—to pilot [11] and more broadly implement the Frailty Management Decision tool. PHNs are not-for-profit organizations funded by the Australian Government to improve access to primary care services for patients, particularly those at risk of poor health outcomes, and coordination of care.

SNHN is a predominantly metropolitan region that is ranked the least socioeconomically disadvantaged PHN in Australia. In contrast, BSPHN is a diverse region that includes metropolitan, rural, and remote island locations, and 2.8% of residents are Aboriginal or Torres Strait Islander. There are 292 general practices in SNHN [19] and 336 in BSPHN [20], with people aged ≥65 years making up 17.3% of SNHN residents and 14.2% of BSPHN residents.

Practices that participated in the pilot program between April 2021 and June 2023 were invited to join this study. Additional practices were recruited between July 2023 and February 2024 via a PHN practice newsletter, expressions of interest, and PHN/researcher practice networks. The Tool was implemented by 24 enrolled practices (14 SNHN, 10 BSPHN).

### Tool Implementation

Patients were screened using the FRAIL Scale tool during the 75+HA. GPs, nurses, and other practice staff completing 75+HA were encouraged to use an optimized health assessment template incorporating the FRAIL Scale questions (Multimedia Appendix 1) or add the FRAIL Scale to their existing practice templates. The frailty score was calculated, and practices referred to the associated management suggestions and interventions recommended within the Tool (Figure 2). Practice staff could download the results of the questionnaire and save them to the patient's medical record.

Training for the practices included (1) an in-person, phone, or web-based video session outlining the aims of the study, how to complete data collection, and use of the FRAIL Scale tool, and (2) check-in visits, or phone/video calls, and emails with the project researcher. Data collection commenced in July 2023 and was completed in January 2025.

### Data Collection

#### *Implementation Evaluation*

During the study, practice staff conducted FRAIL Scale screening with patients aged 75 years and older and collected deidentified data via a REDCap (Research Electronic Data Capture; Vanderbilt University) data collection tool developed for the research. Data collected included date of screening, time to complete the screen, patient characteristics (age, BMI, sex, and Aboriginal or Torres Strait Islander background); whether the patient had been hospitalized (planned or unplanned) or attended the emergency department in the previous 3 months; current medication number; patient FRAIL Scale scores; suggested management/referral options; and any barriers identified by patients or caregivers for accessing suggested management/referral options from a picklist (cost, time, transport, or health) or a free text option.

Practices provided data on the active practice population and the total number of GPs and nurses working in the practice. Practices were classified on location (ie, metropolitan or regional) using the Modified Monash Model [21].

#### *Staff Survey*

Staff perceptions of the long-term sustainment (continued use) of the FRAIL Scale were assessed using the Provider Report



XSL•FO
RenderX

of Sustainment Scale (PRESS), a pragmatic, reliable, valid 3-item survey [22]. The 3 items were rated on a scale from 0 (not at all) to 4 (to a very great extent), with the mean score calculated and higher scores indicating more sustainment. An additional free-text option was included: "Do you have any other feedback regarding the implementation of the FRAIL Scale tool in your practice?"

### *Qualitative*

Qualitative interviews were conducted with (1) practice staff who had used the FRAIL Scale tool during implementation, and (2) patients who had been assessed for risk of frailty to understand determinants (barriers and enablers) to (1) implementing the FRAIL Scale tool and (2) risk of frailty management in primary care.

Staff conducting the FRAIL Scale tool assessments during the study were invited to be interviewed, and practice staff recruited potential patients for interview at the time of the 75+HA. Invitations were emailed or mailed to staff and patient participants, with information describing the rationale for the research provided via an information sheet. Participants were sent the interview questions in advance, allowing them time to consider their viewpoints. Interviews were semistructured and informed by the CFIR [18,23]. The relevance of each CFIR determinant to the implementation of the FRAIL Scale tool was rated by the research team members (CN, JJ, and DC), and those with the highest rating were used to determine interview questions [24]. Interview questions Multimedia Appendix 2) were developed using the Interview Guide Tool available via the CFIR website [25] and with a focus on patient views on the FRAIL Scale assessment tool, barriers to accessing management options and resources, and preferences for suitable options and resources. All interviews were conducted via telephone by an implementation scientist with extensive experience in qualitative research (JJ). As an alternative to data saturation, the process of Malterud et al [26] was used to inform the sample size. Sample size was determined to be met once the purpose and goals of the analysis were achieved, with information power considered strong, since the interviewer was experienced, the research questions were specific, and the interview guide was informed by theory [26]. The interviewer recorded notes during and after interviews to identify barriers and enablers to implementing screening and management for risk of frailty/frailty (JJ). Interviews were audio-recorded, transcribed verbatim using Microsoft Word, cross-checked against the recording (JJ), and anonymized.

## Data Analysis

### *FRAIL Scale*

Data were summarized using median (IQR) for continuous nonnormally distributed variables and frequency (percent) for categorical variables. The associations between frailty (frail, prefrail, and robust) and patient demographics were explored using chi-square tests for categorical characteristics and Mann-Whitney *U* tests for continuous nonnormally distributed characteristics. Statistical significance was indicated at a *P* value of <.05 (2-sided). These analyses were performed in STATA (version 18; StataCorp LLC) and IBM SPSS Statistics (version 29.0.0.0 241).

### *PRESS Survey*

Survey data were presented using descriptive statistics (median and IQR) and tabulated with free-text responses.

### *Qualitative interviews*

A deductive, thematic analysis was used to understand barriers and enablers to implementation of frailty screening and management in general practice with a view to guiding ongoing implementation [27]. Interview transcripts were coded by 2 researchers (JJ and RS), using a deductive process guided by CFIR [18], and themes were developed (JJ and RS). The aim of having 2 coders was not to reach consensus, but to develop richer and more complex insights into the data. As every participant has their own experience and view of frailty screening and management, member checking of results was not conducted [28].

The qualitative feedback was used to explore, in greater depth, the (1) results of the PRESS survey on sustainment of the Tool and (2) data collected by GPs and nurses from patients regarding the barriers they identified to accessing frailty management resources [29].

# Results

## Overview

The characteristics of the Sydney North and Brisbane South practices that implemented the FRAIL Scale tool at the time of the 75+HA are presented in Table 1. All practices were in a metropolitan region (MMM1) [21]. Screening was conducted by 61 staff (57 nurses, 3 GPs, and 1 nondispensing pharmacist) from the practices who used the FRAIL Scale tool.



XSL•FO
RenderX

**Table .** Practice demographics.

| | FRAIL[a] Scale implementation practices, median (IQR) | | |
|---|---|---|---|
| | All (n=23) | BSPHN[b] (n=9)[c] | SNHN[d] (n=14) |
| Practice patient population | 6207.0 (3906.0 - 8477.0) | 7194 (5623.0 - 1999.5) | 4167.5 (2945.8 - 6779.3) |
| Number of active patients aged ≥75 years | 429.0 (209-878) | 461.0 (429.0 - 1272.0) | 348.0 (176.5 - 597.5) |
| Number of GPs[e] in the practice | 7.0 (5.0 - 11.0) | 8.0 (5.5 - 11.5) | 6.5 (4.8 - 11.3) |
| Number of practice nurses | 2.0 (2.0 - 4.0) | 2.0 (1.4 - 4.3) | 2.5 (2.0 - 4.0) |

[a]FRAIL: Fatigue, Resistance, Ambulation, Illnesses, Loss of weight.

[b]BSPHN: Brisbane South Primary Health Network.

[c]Missing data for one BSPHN practice.

[d]SNHN: Sydney North Health Network.

[e]GP: general practitioner.

## Frailty Screening Results

Practice staff screened 1484 patients for risk of frailty at the time of the 75+HA (Table 2). The median time taken by staff to complete the FRAIL Scale tool was 3.9 minutes (IQR 2.2 - 7.6 minutes, n=1458).

**Table .** Patient demographics for Brisbane South and Sydney North primary health networks.

| | Total (n=1484) | Brisbane South (n=588) | Sydney North (n=896) |
|---|---|---|---|
| Patient sex, n (%) | | | |
| Male | 650 (43.8) | 256 (43.5) | 394 (44.0) |
| Female | 834 (56.2) | 332 (56.5) | 502 (56.0) |
| Age in years at screen (years), median (IQR) | 80 (77-85) | 79 (77-84) | 81 (77-85) |
| BMI (kg/m$^2$), median (IQR)[a] | 26.8 (23.7-30) | 27.6 (24.4-30.8) | 26.2 (23.3-29.6) |
| Aboriginal or Torres Strait Islander, n (%)[b] | 7 (0) | 2 (0) | 5 (1) |
| Prescribed medication count, median (IQR)[c] | 5 (3-8) | 5 (3-8) | 5 (3-7) |
| Frail score, median (IQR) | 1 (0-2) | 1 (0-2) | 1 (0-2) |
| Nonelective hospitalization in prior 3 months, n (%)[d] | 56 (4) | 20 (4) | 36 (4) |
| Emergency department attendance in prior 3 months, n (%)[e] | 90 (6) | 38 (6) | 52 (6) |

[a]Missing data, n=271

[b]Missing data, n=40.

[c]Missing data, n=5.

[d]Missing data, n=26.

[e]Missing data, n=5.

At the time of screening, 43.5% (n=645) of all patients (n=1484) were classified as robust (score 0), 41.5% (n=616) as prefrail (scores 1 - 2), and 15.0% (n=223) as frail (scores 3 - 5) (Table 3). People who were frail were more likely to be female, older, have more prescribed medications, and to have had a nonelective hospitalization in the prior 3 months. Although those who were robust had a lower BMI than those who were prefrail or frail (statistically significant), the median BMI did not differ greatly between the frail groups.



XSL•FO
RenderX

**Table .** Patient characteristics by FRAIL (Fatigue, Resistance, Ambulation, Illnesses, Loss of weight) Scale score.

| | Frail score | | | *P* value |
|---|---|---|---|---|
| | Robust (0) (N=645) | Prefrail (1-2) (N=616) | Frail (≥3) (N=223) | |
| Patient sex, n (%) | | | | .045 |
| Male | 299 (46.4) | 269 (43.7) | 82 (37) | |
| Female | 346 (53.6) | 347 (56.3) | 141 (63.2) | |
| Age in years at screen (years), median (IQR) | 79 (77-83) | 80 (77-85) | 84 (79-88) | <.001 |
| BMI (kg/m$^2$), median (IQR)[a] | 26.2 (23.5-29.1) | 27.3 (24-30.8) | 27.1 (23.5-31.2) | <.001 |
| Prescribed medication count, median (IQR)[b] | 4 (2-5) | 6 (4-9) | 8 (5-11) | <.001 |
| Nonelective hospitalization in prior 3 months, n (%)[c] | 5 (1) | 28 (5) | 23 (11) | .01 |
| Emergency department in prior 3 months, n (%)[d] | 20 (3) | 42 (7) | 28 (13) | .82 |

[a]Missing data, n=271.

[b]Missing data, n=5.

[c]Missing data, n=26.

[d]Missing data, n=5.

The predominant frailty management or resources suggested are outlined in Table 4, with GP medication reviews most frequently recommended (17.3%).

**Table .** Management recommendations and resources (10 most frequent for prefrail and frail patients who had recommendations made of the 1484 screened).

| Recommendation or resource | Values, n (%) |
|---|---|
| GP[a] medication review (medication indications/side effects/usage) | 256 (17.3) |
| My Aged Care packages | 194 (13.1) |
| Use depression assessment in 75+HA[b] tool | 141 (9.5) |
| Activity program/ exercise classes | 108 (7.3) |
| Fatigue - screen for sleep apnea, depression, anemia, hypotension, hypothyroidism, B12 deficiency | 107 (7.2) |
| Exercise program with resistance/strength component | 101 (6.8) |
| Support services | 97 (7) |
| Physiotherapy or exercise physiologist for exercise prescription | 95 (6) |
| Pharmacist for comprehensive medication review (Home Medication Review item 900) | 95 (6) |
| Occupational therapy for functional/home review | 66 (4) |

[a]GP: general practitioner.

[b]75+HA: health assessment for those aged 75 years and older.

## PRESS Survey

The PRESS survey was completed by 15 staff (13 nurses, 1 GP, and 1 Practice Manager) from 11 practices (Table 5). All practices had used the FRAIL Scale tool for greater than 6 months prior to survey completion. Free-text comments to the request for other feedback regarding the implementation of the FRAIL Scale tool in the practice are outlined in Multimedia Appendix 3 .



XSL•FO
RenderX

**Table .** PRESS survey feedback (n=15) regarding long-term sustainment (continued use) of the FRAIL (Fatigue, Resistance, Ambulation, Illnesses, Loss of weight) Scale rated on a scale from 0 (not at all) to 4 (to a very great extent).

| Feedback | Median (IQR) |
|---|---|
| Staff use the FRAIL Scale tool as much as possible when appropriate | 3.0 (2.0 - 3.0) |
| Staff continue to use the FRAIL Scale tool throughout changing circumstances | 2.0 (1.0 - 3.0) |
| The FRAIL Scale tool is a routine part of our practice | 2.0 (1.5 - 4.0) |

## Qualitative Interviews

### *Overview*

Of the 16 people who consented and were sent the interview questions, 15 took part in the interviews (94% participation rate): 4 practice nurses, 1 GP, and 10 patients (aged 75 years and older). The qualitative results related to the determinants of implementing the two components of the Tool: (1) frailty screening at the time of the 75+HA and (2) risk of frailty/ frailty management.

### *Determinants of Implementing Frailty Screening at the Time of the 75+HA*

#### Overview

Qualitative feedback highlighted that the FRAIL Scale Tool focused patient and GP attention on aspects of health and aging. Staff identified that screening at the time of the 75 and over health assessment fit with the general practice workflows, and that the FRAIL Scale tool was easy to use, formalized screening, and allowed tracking of frailty over time. However, the PRESS survey comments suggested key factors would support the adoption of the Tool in the long term.

#### The FRAIL Scale Assessment is Recipient-Centered

Patients reported that they appreciated the opportunity to reflect on their health.

> *nobody asks you questions .. like that, usually it's just whatever you're going to the doctor (for) ….the fact that you know at my age 75…I'm still doing well… So I was kind of quite happy with my assessment.* [PS14 (Pt)]
>
> *I think it's great that they do ask those questions because a lot of times you don't mention those sort of things to your doctor when you're there.… those sort of questions .. remind you you're not 21 anymore and your body is not as agile as what it used to be.* [PS22 (Pt)]
>
> *I probably was pleased with what they went through and thought that the questions give me assurance that I'm doing the right thing.* [PS15 (Pt)]
>
> *These questions just remind me that …. I have to pay more attention on my … my daily life. And maybe if no one asked me this, maybe something happening that I didn't notice.* [PS27 (Pt)]
>
> *it's good learning for me …feeling I have to learn more to keep my body healthy.* [PS28 (Pt)]

#### Embedding the FRAIL Scale Tool in Practice Processes

GPs and practice nurses reflected on the benefits of having a process for screening for risk of frailty and a method to record and track risk of frailty in the patient chart over time.

> *I think.. often there's an insight to the doctor because we don't always ask those questions.* [PS17 (GP)]
>
> *they can get an actual number score. Says how frail someone is. If you can look at someone and you can think about their history and everything and they're like, OK, they might be frail. They're becoming a bit frail. But having a number to it makes a difference.* [PS20 (Nurse)]

The FRAIL Scale tool is easy to use, and although it takes extra time to complete, the Tool improves the care for patients.

> *I don't find using the tool complicated at all. I .. it's easy for us to access. It's very self-explanatory.* [PS21 (Nurse)]
>
> *It takes a bit of extra time.. but it means .. we can provide better care for that person… so we have allowed an extra 15 minutes into our appointment.* [PS20 (Nurse)]
>
> *it's worthwhile for the government to do this because you get … better outcome(s)..and the frail elderly, definitely you would save a lot more in preventative care.* [PS17 (GP)]

Staff felt the FRAIL Scale screen fit with the workflows for the 75+HA, and some practices incorporated a link to the Tool from the 75+HA templates in the medical practice software (Best Practice).

> *it's part of the auto fill for the 75+HA. So we do .. daily living and then we go through .. risk factors for osteoporosis. And then falls risk and (the) Frail Scale Tool so it works in perfectly.* [PS18 (Nurse)]

Introduction of the FRAIL Scale tool, which included management recommendation prompts based on the score in each FRAIL domain, was seen as an advantage by staff.

> *much easier because it … gives you prompts … we're actually. .. suggesting actions, whereas before we might have just said that they were frail and .. left it at that, whereas now we're sort of discussing with them .., maybe you should be thinking about doing some balance and falls programs.* [PS18 (Nurse)]

Some staff found the paper-based FRAIL Scale tool useful as a quick guide to screening and management.



XSL•FO
RenderX

> *the frailty management decision tool is quite good.. it's got the words FRAIL down the side and then it's got intervention and referral follow up prompts for things that you can do for each section and that's quite useful. .. easier to look at as an overview.* [PS18 (Nurse)]

> *I do print off recommendations for (management) and (say to the patient) if you want to, you can have a read. You don't have to. I'm not forcing you to do anything, but this information is here for you when you want it and let us know if you want us to refer (you) to anyone.* [PS20 (Nurse)]

With the current format of the Tool, documentation and communication within the practice is part of the process of implementing the Tool.

> *it's in the patient notes.…… this is what the score number was, .. what it meant. And we also save… the whole thing with the recommendations and everything into their notes as well into correspondence* [PS20 (Nurse)]

> *I do find it is a great resource because..then I'll document it in ..the patient notes. But I will also hand it over then to the GP … they will always see the GP after they see me* [PS21 (Nurse)]

However, PRESS survey feedback (Multimedia Appendix 3) highlighted that use of the Tool is not business as usual, and building the FRAIL Scale Tool into the practice management software would automate the recording process and make the Tool "more user friendly".

Patients identified trust in the medical team that facilitated the implementation of screening.

> *[Dr] has been my GP for quite some time and she looks after me and my wife really well. So whatever she asks, I'm very happy to discuss with her.* [PS27 (Pt)]

Some staff felt that completing the risk of frailty screening with robust patients was unnecessary.

> *if they're doing.. 3-4 hours of physical activity a week, you possibly don't need to carry on filling the questionnaire.* [PS21 (Nurse)]

## Determinants of Implementing Risk of Frailty/Frailty Management

### *Overview*

While the FRAIL Scale tool raised awareness of the importance of healthy aging, patients identified barriers to accessing management resources. Access to frailty management interventions and resources was limited due to patient motivation and capability related to health, and limited opportunity due to transport issues, cost of resources, or caring duties. Accessing web-based resources was identified as a challenge for those who may not be savvy with technology, and in this case, face-to-face programs were the preferred option. Data recorded by practice staff at the time of frailty screening is presented in Table 6, with further context provided by quotes from the qualitative interviews.



XSL•FO
RenderX

**Table .** Barriers to accessing frailty management and resources: quantitative (673 prefrail/frail patients who had a recommendation made) and qualitative (10 patients, 1 general practitioner [GP] interview).

| Barriers | Quantitative, n (%) | Qualitative |
|---|---|---|
| Health | 81 (12) | • *I've got blocked arteries in both legs…. I could go out for a walk right now, but I get a certain distance, and I'd have to stop and wait till the… cramps and things in my legs so I can walk home* [PS14 (Pt)] |
| Transport | 81 (12) | • *at over 75, .. it's very difficult for them to access the care either because cost or no one's taking them. They don't want to trouble their family, and so they often don't want to follow through* [PS17 (GP)] |
| Cost | 57 (8) | • *at over 75, .. it's very difficult for them to access the care either because cost or no one's taking them. They don't want to trouble their family, and so they often don't want to follow through* [PS17 (GP)] |
| Time | 57 (8) | • *You gotta have the time and the desire to go walking. So I don't do it too much.* [PS16 (Pt)] |
| Other (caregiver duties, language, grieving, motivation, technology) | 53 (8) | • *I've been with the gym for about 15 years, but .. these days I only….go there every once a week… because of my wife, she has dementia... so I don't have much time to do that* [PS27 (Pt)]<br>• *I'm here by myself… so you sometimes get a little bit lazy* [PS15 (Pt)]<br>• *I haven't been doing them (exercises for shoulder from physio). I've been very lazy, it's been too cold* [PS19 (Pt)]<br>• *another barrier is because we are not IT people. So sometimes ..the programme .. from online .. is a bit difficult for us to join. But if the community can have more practical programmes like face to face .. led by the leader that will be… easy for us to join* [PS28 (Pt)] |

The interviews with patients highlighted other determinants of frailty management in primary care that were not identified by practice staff at the time of frailty screening. These included (1) lack of patient knowledge of management options and resources available, (2) social support enhancing participation in physical activity, and (3) location of resources enhancing access.

### *Knowledge of Available Resources*

Some patients were unaware of resources available locally or were interested in learning more about healthy lifestyle behaviors.

> *They did suggest that maybe that I should… see an exercise Physiologist which .. I've never even heard of before.* [PS26 (Pt)]
>
> *Teach me… the benefit of .. food and how to eat it better. And how to keep the body .., easier to go to the toilet every day ... That would be helpful.* [PS28 (Pt)]

### *Relationships and Social Connectedness Were a Facilitator to Exercise*

Family, friends, and social interaction were identified as supporting physical activity for several people.

> *I go to the gym. I walk with my son every weekend. We have coffee and sometimes, if my daughter can, I walk with her. I see the family all the time. I'm in a good place.* [PS25 (Pt)]
>
> *My daughter gets onto me about that "do your exercises", and my other daughter says the same thing.* [PS19 (Pt)]
>
> *I do Tai chi three times a week…. go to a park with a group… After the Tai Chi we have coffee time and chat.* [PS28 (Pt)]



XSL•FO
RenderX

### *Access to Resources and Local Programs, and Motivation to Attend*

Patients and providers recognized the benefits of local resources for providing access to frailty management.

> *In the retirement village where I am, on Tuesday we have exercises, Wednesday we have a bus which takes us shopping, Thursday I play indoor bowls…. And then Friday night I have tea with the neighbours next door…We have a nursing section in the village and they have.. a gym up there* [PS15 (Pt)]

## Discussion

### Principal Findings

Incorporating the FRAIL Scale tool into the annual health assessment identified 56.5% (839/1484) of the 75 years and over general practice population in this study as frail/at risk of frailty. The predominant frailty management interventions recommended were medication reviews, aged care packages, assessments for depression, and exercise programs. Compared with those who were robust/prefrail, those identified as frail were 4 times more likely to have had a nonelective hospitalization in the 3 months prior to the assessment. Older people living with frailty occupy a large proportion of hospital bed days, at a significant health care cost, and have an increased risk of death [30]. Implementing the FRAIL Scale tool with the annual health assessment provides a funded opportunity to screen for frailty in general practice, prior to onset, and enact prevention/lifestyle interventions to slow or prevent development. Patient feedback highlighted that being involved in the screening process provided an opportunity to reflect on their health, raised awareness of frailty, and prevention and management strategies. Ease of use with screening results linked to actionable management interventions was identified by staff as a valuable component of the FRAIL Scale tool. Patients expressed trust in the primary care team, which may facilitate implementation and highlight the benefits of screening where people receive their usual health care. However, this research has identified individual and health system enablers and barriers experienced by patients to accessing support and services required for effective frailty prevention and management. Determinants of frailty management related to the availability of local resources and the capability, opportunity, and motivation for patients to access these resources. Increasing knowledge for patients and providers on frailty management, providing social support for healthy aging activities, and improving access to management resources were highlighted as potential facilitators to healthy aging.

### Comparison to Prior Work

The rates of prefrailty and frailty (839/1484, 56.5%) in this study are lower than those identified in the pilot study (860/1071, 80%), where both opportunistic screening (at the time of routine procedures) and structured screening were conducted with those aged 75 years and older [11]. This suggests routine screening has the potential for early identification of prefrailty and presents an opportunity for health promotion in early aging.

Other researchers, however, have identified that primary care frailty screening and management tools must have usability and fit with practice workflows and have highlighted the importance of resources or funding for the team to complete frailty assessment [31-34]. This study supports previous findings that implementing the FRAIL Scale tool at the time of the Medicare-funded 75+HA fits with practice workflows, formalizes processes, and renumerates staff time to screen for frailty in this population [11]. In addition, as found in the pilot study, some staff adopted strategies to facilitate the workflow, including incorporating the Tool in the 75+HA templates [11]. However, survey feedback highlighted that embedding the screening tool within practice software was needed to support broader adoption and sustainability of the FRAIL Scale tool in primary care.

In addition, supporting frailty prevention and management for older people in primary care requires knowledge of resources available [15]. While practitioners appreciated the management prompts incorporated in the FRAIL Scale tool, practitioner skill is required to use and encourage behavior change techniques [35] such as social support [36] and goal setting [37] to support healthy aging.

### Strengths

With limited research on prefrailty and frailty in community-dwelling adults in Australia [2,38], this research has highlighted the association of hospitalization with increasing rates of frailty in 2 diverse PHN regions in Australia [39].

The Implementation Research Logic Model [40] has been used throughout the research process to guide planning for the implementation of the FRAIL Scale [41]. The use of the Implementation Research Logic Model has enhanced the implementation process and understanding of implementation determinants framed by CFIR and informed strategies for broader implementation.

### Limitations

First, screening was conducted across only 24 general practices in regions with larger metropolitan populations of older persons and may not be representative of all patients aged 75 years and older. While participation was offered to all practices in each study region, recruitment of metropolitan practices may reflect those practices that had the opportunity to be involved due to adequate staffing, which is often limited in regional, rural, and remote general practices [42]. Preliminary pilot work by the research team in rural and remote Queensland suggests rates of frailty may be higher in nonmetropolitan regions. Second, as risk of frailty screening was completed as part of the 75+HA in primary care settings, the rates of frailty and prefrailty found in this study may not reflect the rates in the broader population of community-dwelling adults aged 75 years or older. Evaluating screening through other avenues, such as community organizations, will be important to reach a broader representation of this age group. Third, the rates of screening for the Aboriginal and Torres Strait Islander population were low in this study. Frailty is known to occur more frequently and at a younger age in the Aboriginal and Torres Strait Islander population [38], and the acceptability of the FRAIL Scale tool to this population



XSL•FO
RenderX

was not assessed [38]. Fourth, only 25% (n=15) of staff completed the PRESS survey, with just under 50% (n=11) of practices represented, and qualitative interviews were conducted with only 5 staff. While these results may not reflect all users of the FRAIL Scale, after considering the results from previous interviews with practice nurses and GPs conducted in the pilot study [11] and the free text feedback from the PRESS survey, we did not feel we were gathering any further information and focused on the patient interviews. Finally, while practice staff recorded the management recommendations provided to patients based on the Tool, data was not collected on recommendations adopted by patients following their appointment. In addition, qualitative interviews with patients indicated that social relationships are a facilitator to healthy aging, yet social connectedness was not considered in the FRAIL Scale.

### Future Research

Screening and management with the Tool is currently being evaluated in Western Queensland, where 20% of the population has an Aboriginal or Torres Strait Islander background [43], to understand rates of frailty and the appropriateness of the Tool for other populations and in rural and remote settings. Screening and management recommendations for addressing social connectedness have been incorporated into the Tool for this ongoing research.

While the FRAIL Scale tool can be readily incorporated into the 75+HA, broader implementation requires the Tool to be incorporated into practice software systems. Endorsement of FHIR (Fast Healthcare Interoperability Resources) standards by the Australian Digital Health Agency [44] has provided an opportunity for the research team to develop and evaluate a FHIR-compatible prototype of the FRAIL Scale tool in conjunction with CSIRO researchers [45]. This will improve interoperability across health systems and allow appropriate sharing of data between practice nurses, GPs, and health providers in other health care settings [46].

Importantly, further research is required to understand the effect of screening for risk of frailty in general practice on frailty awareness in the population and whether management recommendations increase adoption of health behaviors to prevent and reverse frailty, encourage proactive self-management, and slow or reverse frailty progression [47]. Consumer engagement will be vital in the development of appropriate, accessible resources for promoting the benefits of frailty prevention and management [7].

While NSPHN and BSPHN are relatively well-resourced regions with accessible programs and support networks, many regions of Australia lack the accessible, coordinated, multidisciplinary workforce required to effectively identify and manage the risk of frailty. This is made more difficult by the poor integration of Commonwealth, state, private, and not-for-profit services [48,49], identified in the aged care sector [50], and exacerbated in rural Australia by a limited local workforce to deliver effective care [42]. Frailty management would benefit from an alliance governance approach, an internationally accepted methodology, effective in binding such diverse health organizations contractually to collectively deliver an agreed program of work [51]. The research team is currently evaluating the implementation of an alliance governance model to support healthy aging in Western Queensland.

### Conclusions

Screening and management of frailty and risk of frailty are not routine in general practice in Australia. A structured approach, using the FRAIL Scale tool at the time of the annual health assessment, has identified 56.5% of the general practice population aged 75 years and older as frail/at risk of frailty. With higher rates of nonelective hospitalization associated with those identified as frail compared with those who were robust/prefrail, routine screening that fits with practice workflows is important in the primary care setting. Importantly, the FRAIL Scale tool provides staff with intervention options to support healthy aging.

### Acknowledgments

We would like to acknowledge the contribution of the partner organizations, the Northern Sydney Primary Health Network trading as Sydney North Health Network (SNHN) and Brisbane South Primary Health Network (BSPHN), and the general practices that conducted the data collection.

### Funding

This research was funded by a Translational Research Grant from the HCF Research Foundation.

### Data Availability

The datasets analyzed for this study are available from the corresponding author on reasonable request.

### Authors' Contributions

JJ, CN, RS, DC, AP, and CJ are the principal and associate investigators of the project. JJ, CN, DC, RS, and CJ were extensively involved in the development of the study concept and design and provided project governance. RS and DC collected the data, and JJ and AP designed and conducted the analysis. JJ, RS, CN, and CJ drafted the manuscript. All authors were involved in critically revising the manuscript and read and approved the final manuscript for submission to your journal.



XSL•FO
RenderX

## Conflicts of Interest

CJ is a director on the HCF Research Foundation Board.

Multimedia Appendix 1
75+ Health Assessment_optimised.
[DOCX File, 60 KB - i-jmr_v15i1e79681_app1.docx ]

Multimedia Appendix 2
Qualitative interview guides_staff and patients.
[DOCX File, 23 KB - i-jmr_v15i1e79681_app2.docx ]

Multimedia Appendix 3
Free text survey responses.
[DOCX File, 15 KB - i-jmr_v15i1e79681_app3.docx ]

## References


1. Ageing and health. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2025-11-15]
2. O'Caoimh R, Sezgin D, O'Donovan MR, et al. Prevalence of frailty in 62 countries across the world: a systematic review and meta-analysis of population-level studies. Age Ageing 2021 Jan 8;50(1):96-104. [doi: 10.1093/ageing/afaa219] [Medline: 33068107]
3. Morley JE, Vellas B, van Kan GA, et al. Frailty consensus: a call to action. J Am Med Dir Assoc 2013 Jun;14(6):392-397. [doi: 10.1016/j.jamda.2013.03.022] [Medline: 23764209]
4. Gill TM, Gahbauer EA, Han L, Allore HG. The relationship between intervening hospitalizations and transitions between frailty states. J Gerontol. A Biol Sci Med. Sci 2011 Nov 1;66A(11):1238-1243. [doi: 10.1093/gerona/glr142]
5. Travers J, Romero-Ortuno R, Bailey J, Cooney MT. Delaying and reversing frailty: a systematic review of primary care interventions. Br J Gen Pract 2019 Jan;69(678):e61-e69. [doi: 10.3399/bjgp18X700241] [Medline: 30510094]
6. Dent E, Morley JE, Cruz-Jentoft AJ, et al. Physical frailty: ICFSR international clinical practice guidelines for identification and management. J Nutr Health Aging 2019 Nov;23(9):771-787. [doi: 10.1007/s12603-019-1273-z]
7. Dent E, Hanlon P, Sim M, et al. Recent developments in frailty identification, management, risk factors and prevention: a narrative review of leading journals in geriatrics and gerontology. Ageing Res Rev 2023 Nov;91:102082. [doi: 10.1016/j.arr.2023.102082] [Medline: 37797723]
8. Job J, et al. Abstracts of the australasian association for academic primary care (AAAPC). Presented at: Australian Journal of Primary Health; Aug 17-18, 2023.
9. Morley JE, Malmstrom TK, Miller DK. A simple frailty questionnaire (FRAIL) predicts outcomes in middle aged African Americans. J Nutr Health Aging 2012 Aug;16(7):601-608. [doi: 10.1007/s12603-012-0084-2]
10. Thompson MQ, Theou O, Tucker GR, Adams RJ, Visvanathan R. FRAIL scale: predictive validity and diagnostic test accuracy. Australas J Ageing 2020 Dec;39(4):e529-e536. [doi: 10.1111/ajag.12829] [Medline: 32748992]
11. Job J, Nicholson C, Clark D, Arapova J, Jackson C. The feasibility, acceptability and appropriateness of screening for frailty in Australians aged 75 years and over attending Australian general practice. Aust J Prim Health 2024 May;30:30. [doi: 10.1071/PY23173] [Medline: 38739739]
12. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol. A Biol Sci Med. Sci. 2001 Mar 1;56(3):M146-M157. [doi: 10.1093/gerona/56.3.M146]
13. Dent E, Lien C, Lim WS, et al. The Asia-Pacific clinical practice guidelines for the management of frailty. J Am Med Dir Assoc 2017 Jul;18(7):564-575. [doi: 10.1016/j.jamda.2017.04.018]
14. Ruiz JG, Dent E, Morley JE, et al. Screening for and managing the person with frailty in primary care: ICFSR consensus guidelines. J Nutr Health Aging 2020;24(9):920-927. [doi: 10.1007/s12603-020-1492-3] [Medline: 33155616]
15. Shafiee Hanjani L, Fox S, Hubbard RE, et al. Frailty knowledge, training and barriers to frailty management: a national cross-sectional survey of health professionals in Australia. Australas J Ageing 2024 Jun;43(2):271-280. [doi: 10.1111/ajag.13232] [Medline: 37563782]
16. Roberts S, Gomes K, Rattray M. Dietitians' perceptions of identifying and managing malnutrition and frailty in the community: a mixed-methods study. Nutr Diet 2023 Nov;80(5):511-520. [doi: 10.1111/1747-0080.12799] [Medline: 36843203]
17. Mitchell EKL, D Amore A. Recent trends in health assessments for older Australians. Aust J Prim Health 2021 Jun;27(3):208-214. [doi: 10.1071/PY20283] [Medline: 34148567]
18. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated consolidated framework for implementation research based on user feedback. Implement Sci 2022 Oct 29;17(1):75. [doi: 10.1186/s13012-022-01245-0] [Medline: 36309746]




XSL•FO
RenderX

19. Needs assessment 2024-2027. : Northern Sydney Primary Health Network; 2024.
20. State of Queensland (Metro South Hospital and Health Service) and Brisbane South PHN, Brisbane South joint regional needs assessment 2025-2027. : Queensland Government, Metro South Health; 2024.
21. Health workforce locator. Australian Government Department of Health and Aged Care. 2023. URL: https://www.health.gov.au/resources/apps-and-tools/health-workforce-locator/app [accessed 2024-12-01]
22. Moullin JC, Sklar M, Ehrhart MG, Green A, Aarons GA. Provider Report of Sustainment Scale (PRESS): development and validation of a brief measure of inner context sustainment. Implement Sci 2021 Aug 30;16(1):86. [doi: 10.1186/s13012-021-01152-w] [Medline: 34461948]
23. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4:50. [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
24. Haverhals LM, Sayre G, Helfrich CD, et al. E-consult implementation: lessons learned using consolidated framework for implementation research. Am J Manag Care 2015 Dec 1;21(12):e640-e647. [Medline: 26760426]
25. Consolidated Framework for Implementation Research (CFIR). 2025. URL: https://cfirguide.org [accessed 2025-11-15]
26. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
27. Hamilton AB, Finley EP. Qualitative methods in implementation research: an introduction. Psychiatry Res 2019 Oct;280:112516. [doi: 10.1016/j.psychres.2019.112516]
28. Motulsky SL. Is member checking the gold standard of quality in qualitative research? Qual Psychol 2021;8(3):389-406. [doi: 10.1037/qup0000215]
29. Plano Clark VL. Meaningful integration within mixed methods studies: identifying why, what, when, and how. Contemp Educ Psychol 2019 Apr;57:106-111. [doi: 10.1016/j.cedpsych.2019.01.007]
30. Hoogendijk EO, Afilalo J, Ensrud KE, Kowal P, Onder G, Fried LP. Frailty: implications for clinical practice and public health. Lancet 2019 Oct 12;394(10206):1365-1375. [doi: 10.1016/S0140-6736(19)31786-6] [Medline: 31609228]
31. Ng YX, Cheng LJ, Quek YY, Yu R, Wu XV. The measurement properties and feasibility of FRAIL scale in older adults: a systematic review and meta-analysis. Ageing Res Rev 2024 Mar;95:102243. [doi: 10.1016/j.arr.2024.102243] [Medline: 38395198]
32. Chicoulaa B, Escourrou E, Durrieu F, et al. Challenges in management of frailty by primary healthcare teams: from identification to follow-up. La Presse Médicale Open 2022;3:100032. [doi: 10.1016/j.lpmope.2022.100032]
33. Lawson B, Sampalli T, Warner G, et al. Improving care for the frail in Nova Scotia: an implementation evaluation of a frailty portal in primary care practice. Int J Health Policy Manag 2019 Feb 1;8(2):112-123. [doi: 10.15171/ijhpm.2018.102] [Medline: 30980624]
34. Warner G, Lawson B, Sampalli T, Burge F, Gibson R, Wood S. Applying the consolidated framework for implementation research to identify barriers affecting implementation of an online frailty tool into primary health care: a qualitative study. BMC Health Serv Res 2018 May 31;18(1):395. [doi: 10.1186/s12913-018-3163-1] [Medline: 29855306]
35. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
36. Lindsay Smith G, Banting L, Eime R, O'Sullivan G, van Uffelen JGZ. The association between social support and physical activity in older adults: a systematic review. Int J Behav Nutr Phys Act 2017 Apr 27;14(1):56. [doi: 10.1186/s12966-017-0509-8] [Medline: 28449673]
37. Salih SA, Koo A, Boland N, Reid N. Attainment of community-based goals is associated with lower risk of hospital readmission for older Australians accessing the Australian transition care program. Int J Environ Res Public Health 2025 Jul 22;22(8):1162. [doi: 10.3390/ijerph22081162] [Medline: 40869749]
38. Lewis ET, Howard L, Carroll UR, et al. Looking out across the front yard: aboriginal peoples' views of frailty in the community – a qualitative study. Ethn Health 2024 Nov 16;29(8):987-1007. [doi: 10.1080/13557858.2024.2402527]
39. Dent E, Dal Grande E, Price K, Taylor AW. Frailty and usage of health care systems: results from the South Australian monitoring and surveillance system (SAMSS). Maturitas 2017 Oct;104:36-43. [doi: 10.1016/j.maturitas.2017.07.003] [Medline: 28923175]
40. Smith JD, Li DH, Rafferty MR. The implementation research logic model: a method for planning, executing, reporting, and synthesizing implementation projects. Implement Sci 2020 Dec;15(1):84. [doi: 10.1186/s13012-020-01041-8]
41. Moullin JC, Dickson KS, Stadnick NA, et al. Ten recommendations for using implementation frameworks in research and practice. Implement Sci Commun 2020;1(1):42. [doi: 10.1186/s43058-020-00023-7] [Medline: 32885199]
42. Mullan L, Armstrong K, Job J. Barriers and enablers to structured care delivery in Australian rural primary care. Aust J Rural Health 2023 Jun;31(3):361-384. [doi: 10.1111/ajr.12963] [Medline: 36639909]
43. Western Queensland PHN, Western Queensland PHN fact sheet. : Australian Government, Department of Health, Disability and Ageing; 2024.
44. Connecting Australian healthcare national healthcare interoperability plan 2023–2028. : Australian Government, Australian Digital Health Agency; 2023.




XSL•FO
RenderX

45. Data standards for interoperability. : CSIRO Australia's National Science Agency; 2023.
46. Interoperability and digital health standards. Australian Digitial Health Agency. 2025. URL: https://developer.digitalhealth.gov.au/initiatives/interoperability-and-digital-health-standards [accessed 2025-11-15]
47. Zheng L, Li G, Qiu Y, Wang C, Wang C, Chen L. Clinical practice guidelines for the prevention and management of frailty: a systematic review. J Adv Nurs 2022 Mar;78(3):709-721. [doi: 10.1111/jan.15067] [Medline: 34617618]
48. Jackson AM C. Big talk, little action: the enduring narrative of primary care reform. Aust Health Rev 2024 Jun 6;48(3):219-221. [doi: 10.1071/AH24090]
49. Jackson CL, O'Halloran D. Reforming our health care system: time to rip off the band-aid? Med J Aust 2021 Oct 4;215(7):301-303. [doi: 10.5694/mja2.51261] [Medline: 34519047]
50. Royal Commission into Aged Care Quality and Safety: Commonwealth of Australia; 2021, Vol. 2.
51. Nicholson C, Hepworth J, Burridge L, Marley J, Jackson C. Translating the elements of health governance for integrated care from theory to practice: a case study approach. Int J Integr Care 2018 Jan 31;18(1):11. [doi: 10.5334/ijic.3106] [Medline: 29588645]

## Abbreviations

**75+HA:** health assessment for those aged 75 years and older
**BSPHN:** Brisbane South Primary Health Network
**CFIR:** Consolidated Framework for Implementation Research
**FHIR:** Fast Healthcare Interoperability Resources
**FRAIL:** Fatigue, Resistance, Ambulation, Illnesses, Loss of weight
**GP:** general practitioner
**PHN:** primary health network
**PRESS:** Provider Report of Sustainment Scale
**REDCap:** Research Electronic Data Capture
**SNHN:** Sydney North Health Network

*Edited by A Schwartz, M Balcarras, N Cahill; submitted 25.Jun.2025; peer-reviewed by R Ogrin, SQ Yoong; revised version received 27.Nov.2025; accepted 29.Dec.2025; published 02.Mar.2026.*

*Please cite as:*
*Job JR, Nicholson C, Strauss R, Clark D, Pelecanos A, Jackson C*
*Frailty Screening and Management for Older Australians in General Practice: Mixed Methods Evaluation*
*Interact J Med Res 2026;15:e79681*
*URL: https://www.i-jmr.org/2026/1/e79681*
*doi:10.2196/79681*

© Jennifer R Job, Caroline Nicholson, Ruby Strauss, Debra Clark, Anita Pelecanos, Claire Jackson. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 2.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Modeling Frailty Risk Based on Biochemical and Hematological Parameters Using Multiple and Logistic Regression: Cross-Sectional Analysis of the Toledo Study for Healthy Aging

Francisco Zamudio[1], PhD; Diego Arauna[2], PhD; Francisco Garcia[3], PhD; Cecilia Albala[4], PhD; Victoria Villalobos[2], MT; Eduardo Fuentes[2], PhD; Ivan Palomo[2], PhD

[1]Department of Public Health, University of Talca, Talca, Chile

[2]Thrombosis Research Center, Interuniversity Center for Healthy Aging, VITALIS Longevity Center, Medical Technology School, Department of Clinical Biochemistry and Immunohaematology, Faculty of Health Sciences, University of Talca, Avenida Lircay S/N, Talca, Chile

[3]Department of Geriatrics, Hospital Virgen del Valle, Complejo Hospitalario Universitario de Toledo, Toledo, Spain

[4]Instituto de Nutrición y Tecnología de los Alimentos, University of Chile, Santiago, Chile

**Corresponding Author:**
Ivan Palomo, PhD
Thrombosis Research Center, Interuniversity Center for Healthy Aging, VITALIS Longevity Center, Medical Technology School, Department of Clinical Biochemistry and Immunohaematology, Faculty of Health Sciences, University of Talca, Avenida Lircay S/N, Talca, Chile

## Abstract

**Background:** Frailty is a complex syndrome characterized by diminished physiological reserves, resulting in increased vulnerability among older adults. Scalable tools that leverage routinely collected laboratory data could enable earlier risk stratification in aging populations.

**Objective:** The aim of the study is to model frailty risk using Frailty Trait Scale (FTS)-12 scores and biochemical or hematological parameters, and to examine sex-specific patterns in predictors and performance.

**Methods:** This study analyzed data from a population-based cohort of 1485 individuals aged 65 years or older, drawn from the Toledo Study for Healthy Aging (TSHA). Multiple and logistic regression analyses were used to investigate the relationships between biochemical and hematological parameters, and the FTS-12 score. Predictive models were developed separately for female and male participants, incorporating both single and composite risk factors. Model performance was evaluated using correlation coefficients, coefficients of determination, Akaike information criterion, log-likelihood values, and receiver operating characteristic curves, which summarize how well the models distinguish between higher and lower frailty risk.

**Results:** The regression models accounted for approximately 40% of the variability in FTS-12 scores, revealing distinct risk factors for each sex. Gamma-glutamyl transferase, albumin, and age were kept as key predictors in the selected models. Several composite risk factor functions were also significantly associated with the FTS-12 score. Correlations between observed and predicted FTS-12 scores were moderately strong, with values of 0.621 ($P$<.001) for female participants and 0.615 ($P$<.001) for male participants. Logistic regression models showed moderate-to-good discriminative performance for high frailty probability, achieving receiver operating characteristic area under the curve values of 0.799 for female participants and 0.824 for male participants.

**Conclusions:** This analysis is an exploratory proof-of-concept study, and the developed models do not constitute validated clinical prediction tools. Routine laboratory profiles demonstrate moderate, sex-specific predictive performance for FTS-12–defined frailty and may, in principle, complement comprehensive geriatric assessment by facilitating the early identification of patients who could benefit from further evaluation and closer clinical follow-up. External validation in independent cohorts is required before any clinical use, alongside longitudinal designs with repeated biomarker measures and integration of additional biological layers, to enhance generalizability and confirm clinical utility. Within these limits, the analysis highlights the potential of routine laboratory data to support digitally enabled frailty screening and workflow-integrated risk stratification in primary care and remote geriatric care settings.

***(Interact J Med Res 2026;15:e76500)*** doi:10.2196/76500

**KEYWORDS**

aging; frailty; frailty risk prediction; hematological parameters; biochemical parameters; logistic regression analysis



XSL•FO
RenderX

## Introduction

Frailty is a multifaceted clinical syndrome characterized by diminished strength, endurance, and physiological function, leading to increased vulnerability to dependency and mortality [1-3]. Its complexity stems from several subtypes, including physical, cognitive, oral, and social frailty [4-6]. Physical frailty and physiological decline have been the most extensively studied. Fried et al [3] distinguished frailty as a condition separate from aging, comorbidity, and disability, noting its clinical reversibility and its role as a precursor to dependency in older adults. While unhealthy aging provides a context for frailty development, it is not an inevitable outcome, as individuals may exhibit varying levels of frailty over time [1,7-9]

Assessment tools such as the Fried Frailty Phenotype and the Frailty Index (FI) are routinely used in clinical practice to evaluate frailty. The Fried Frailty Phenotype identifies 5 dimensions—weakness, low endurance, slowness, low physical activity, and weight loss—and classifies individuals as frail if they perform poorly in 3 or more domains [3]. In contrast, the FI conceptualizes frailty as the accumulation of deficits that predispose individuals to adverse outcomes, including mortality and hospitalization [10]. Despite their widespread validation, both methods have limitations in clinical settings, such as being time-intensive and challenging for monitoring subtle changes over time [11,12]. To address these gaps, newer approaches have expanded the conceptualization of frailty. García-García et al [11,13] introduced the Frailty Trait Scale (FTS)-12, which considers frailty as a biological trait encompassing additional domains, including vascular and nervous system health and nutritional status. The FTS has 3 validated versions (FTS-12, FTS-5, and FTS-3), each capturing different dimensions of frailty. These scales exhibit high sensitivity and accuracy in detecting frailty and predicting adverse health outcomes, such as hospitalization, fractures, and mortality. Among these, FTS-12 is effective in comprehensively assessing frailty and its associated risks [11].

Since 2022, we have pursued a screening-oriented approach to investigate whether routinely available laboratory data can help identify older adults at increased risk of frailty. In this study, we examined whether common biochemical and hematological parameters could be used to predict the FTS-12 score and to identify individuals more likely to fall within the highest frailty range. Beyond its statistical relevance, this approach may contribute to health care innovation by evaluating whether low-cost, routinely available laboratory data can support practical frailty risk stratification within digitally enabled geriatric care. In this context, this study aims (1) to develop an equation to predict an FTS-12 score of an older individual based on their biochemical profile; (2) to determine whether frailty, as assessed by the FTS-12 scale, can be predicted to fall within the upper range of the distribution using routinely available blood-based biomarkers; and (3) to establish exploratory criteria for selecting predictive equations based on observed frailty estimates.

This study is conceived as an exploratory, proof-of-concept analysis, and the resulting models should not be regarded as validated clinical prediction tools. The approach is not intended to replace comprehensive clinical assessment of frailty; rather, it explores whether routinely available laboratory data carry meaningful frailty-related signals that could, pending external validation, eventually support preliminary screening of older individuals with a high probability of frailty.

## Methods

### Data Origin

This analysis used data from the Toledo Study for Healthy Aging (TSHA), a prospective, population-based cohort study investigating the determinants and consequences of frailty syndrome in older adults living in Toledo, Spain [14]. The study included 1972 participants who completed a home-based interview and physical evaluation, with biological samples provided by 1752 individuals for laboratory analysis. Baseline data were collected between June 2006 and September 2009, and participants were followed for an average of 3.5 (SD 1.19; range 0.30 - 4.79) years to evaluate study outcomes [11].

### FTS-12, Blood Collection, and Measured Traits

The FTS-12 served as the dependent variable in statistical analyses. This scale includes 12 items across 7 dimensions. To calculate the FTS-12 score for each participant, the TSHA collected data in the following areas: (1) self-reported comorbidities, (2) basic and instrumental activities of daily living, (3) depressive symptoms, and (4) cognitive function. Trained nurses conducted physical examinations and clinical tests at participants' homes to supplement these data. Detailed assessment methods are provided in García-García et al [11].

Participants also provided 12-hour fasting blood samples, from which two primary categories of variables were derived: (1) 15 variables assessing blood characteristics (hemogram) and (2) 19 variables measuring biochemical characteristics (biochemical profile), along with an added 12 unique biochemical traits not included in the standard profile. In total, 46 blood-related variables, combined with participants' age, were included in the original database.

### Ethical Considerations

The TSHA was approved by the clinical research ethics committee of the Complejo Hospitalario de Toledo, Toledo, Spain. The original cohort study ("Estudio Toledo de Envejecimiento") received ethics approval on March 30, 2005 (application 22). In addition, the specific research project entitled "Predictors of frailty, disability, cognitive impairment, vascular events, hospitalization, and mortality in a cohort of adults older than 64 years. Toledo Study for Healthy Aging" was approved on July 14, 2010 (application 93). All participants provided written informed consent before enrollment. This study is a secondary analysis of previously collected cohort data.



XSL•FO
RenderX

## Statistical Analysis

### *Exploratory Data Analysis*

The software programs JMP (version 13.2) and SAS Enterprise Guide (version 7.1; SAS Institute Inc) were used to perform descriptive statistical methods for each sex independently and to identify potential data anomalies [15,16]. Descriptive statistics included (1) examining data dispersion for individual variables and correlations between potential explanatory variables, (2) identifying plausible probability distributions for each variable, (3) detecting potential outliers mainly by using box plot analysis, and (4) analyzing the relationship between the main dependent variable (FTS-12) and possible explanatory variables. The final database included measurements from 1485 individuals aged 65 years and older, comprising 848 female participants and 637 male participants.

### *Missing Data*

Missing data were present in less than 5% of observations across study variables. Analyses were conducted using complete cases only (listwise deletion). This approach is valid under the reasonable assumption that data were missing completely at random. Since the small amount of missingness was mostly caused by small problems during blood sample collection, this assumption was deemed reasonable. Given the low proportion of missingness, potential bias due to exclusion of incomplete cases was considered negligible. The exclusion of missing data is the default procedure in SAS Enterprise Guide (version 7.1) software.

### *Modeling Regression Equations to Predict the FTS-12 Score*

Two predictive models of FTS-12 scores were developed based on the following criteria: (1) sex and (2) available biological information, which included data from the biochemical profile and hemogram. The models were constructed using the theory of linear models in the framework of multiple regression analysis [17-19]. The mathematical formulations of the regression models are provided in Multimedia Appendix 1. Additional details regarding the model and its underlying assumptions are provided in Multimedia Appendix 2. We used a 2-stage, data-driven selection process. Predictors were initially screened based on their Pearson correlation with the outcome $r \geq 0.18$. A stepwise multiple regression was then applied to the remaining variables [17,18]. This approach was used for exploratory purposes, and the results should be interpreted cautiously. This approach was selected and justified because (1) we have a very large number of predictors (n=46); (2) this is a proof-of-concept analysis for predictive purposes (FTS-12 score and high-frailty probability, as described below), rather than an inferential analysis; (3) results must be treated as "exploratory and preliminary," and the stepwise was used as a diagnostic tool, not a decision rule; and (4) the final model will be validated on independent data. Selection criteria included the coefficient of determination, residual mean square, Mallows $C_p$ statistic, and Akaike information criterion (AIC). All statistical calculations were performed using the PROC REG procedure in SAS (version 9.4).

Bootstrapping in multiple regression can be used to support statistical inference when classical assumptions are questionable or when sample size is limited. Bootstrapping was unnecessary in our multiple regression analysis. The size of the database is large; residual effects followed a normal distribution, and there were no heteroscedasticity problems. Here, bootstrapping is optional. However, we considered performing bootstrapping as a complement to the selection process. We used the stepwise selection procedure, and bootstrapping was applied to construct empirical CIs for the parameter estimates. The "bootstrapping" analysis was then conducted for the selected multiple regression models using the PROC SURVEYSELECT and PROC REG procedures in SAS (version 9.4). We generated 1000 subsamples of size 500 from the dataset and fitted a regression model for each subsample. From these models, 95% CIs for each parameter and the model's coefficient of determination were computed by calculating the 2.5% and 97.5% percentiles. A summary of bootstrapping is shown in Multimedia Appendix 2. Because the multiple regression model predicts a continuous outcome, calibration in the probabilistic sense was not applicable; instead, model performance was evaluated using goodness-of-fit measures (eg, $R^2$) and residual diagnostics and agreement between observed and predicted values. These approaches provide an assessment of model performance analogous to calibration in probabilistic models.

### *Modeling the Odds of Reaching Maximum FTS-12 Scores*

FTS-12 scores ranged from 1 to 100. To analyze frailty levels, the 3 quartiles of the distribution were calculated, dividing frailty data into 4 distinct quarters [11]. Each individual's FTS-12 score was assigned to one of these 4 quarters, treating an individual's FTS-12 value in any quarter as a binomial (binary) variable. The fourth quarter—standing for the highest frailty levels—was of primary interest in this study. This quarter included FTS-12 values above the third quartile (scores >55), representing the highest frailty scores in this cohort. This upper-quartile definition was used as a cohort-specific analytical threshold to identify the subgroup with the highest frailty burden within the present sample and should not be interpreted as a universal clinical cutoff for the FTS-12 scale. For any individual in the original TSHA data with an FTS-12 value equal to or above 55, a "1" was assigned under a newly created "fourth quarter" column, while scores below this threshold were assigned a "0." We defined a binary response variable, $Y$, such that $Y=1$ if an individual's FTS-12 score exceeded the third quartile (ie, fell within the highest—fourth—quarter), and $Y=0$ otherwise. Therefore, a multiple logistic regression analysis was conducted for the binary variable representing the fourth quarter, following the approach described by Hosmer et al [20]. Logistic regression enables the prediction of the odds of a particular event occurring as a function of observed values for a set of potential explanatory variables. Depending on the context, these variables are referred to as independent variables, risk factors, predictors, or regressors. Additional details regarding the logistic model and its underlying assumptions are provided in Multimedia Appendix 2. Regression analysis was conducted using JMP (version 13.2) software, along with the LOGISTIC procedure in SAS (version 9.4) and SAS Enterprise



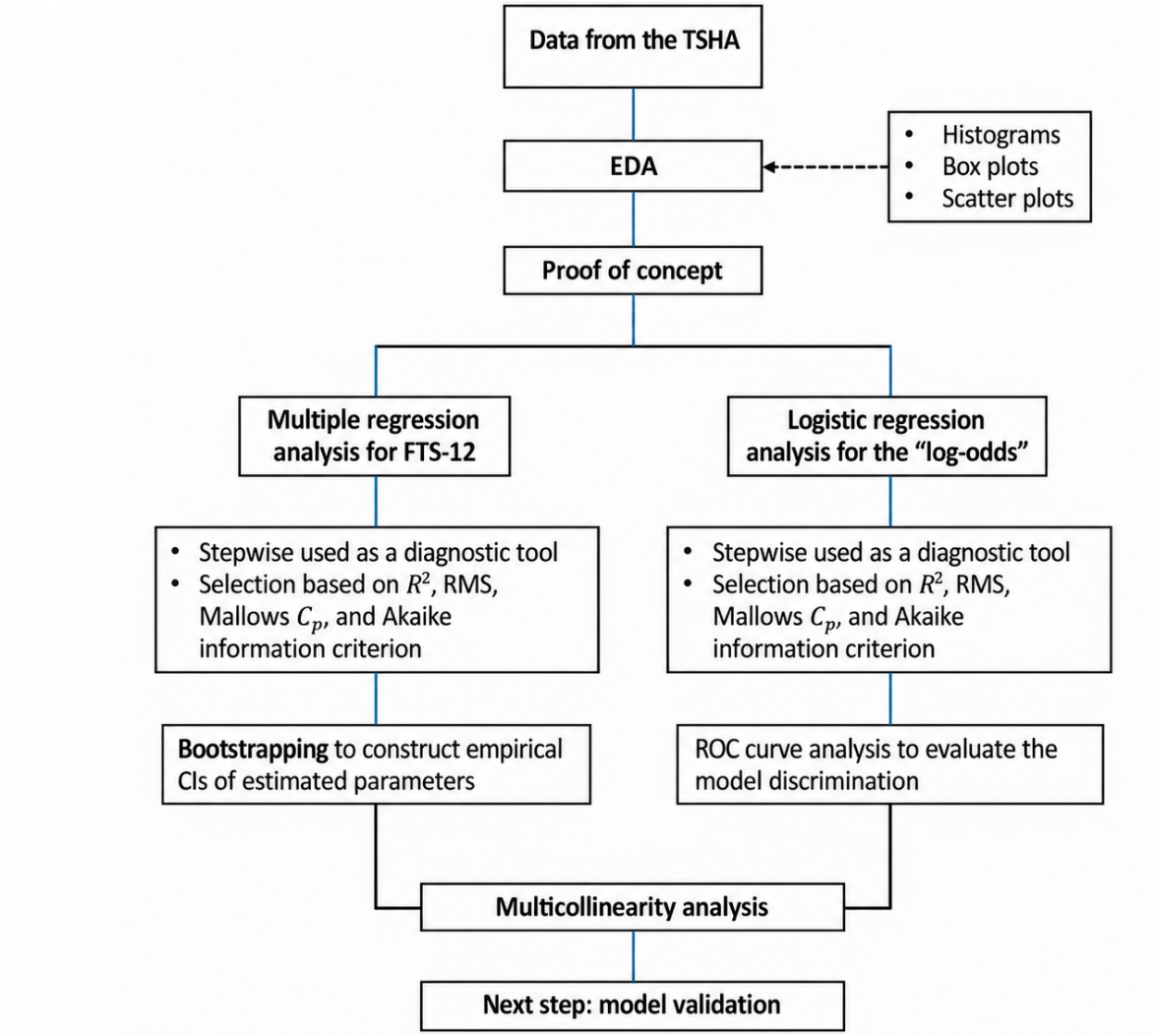
XSL•FO
RenderX

Guide (version 7.1). A stepwise regression approach, as outlined by Hosmer and Lemeshow [20], was applied as the selection method. Selection criteria included the coefficient of determination, residual mean square, Mallows $C_p$ statistic, and AIC. Finally, we generated receiver operating characteristic (ROC) curves to illustrate how well the binary models distinguished between higher and lower frailty probability. As in the multiple regression analysis, this approach was used for exploratory purposes, and the results should be interpreted cautiously. Bootstrapping was not performed in the logistic regression analysis because of the large sample size and the fact that the number of events per variable exceeded 20 in the binary outcome. Instead, ROC curve analysis was used to assess how well the models distinguished between higher and lower frailty probability. This analysis provided area under the curve (AUC) estimates for the selected variables and allowed comparison of their discriminatory performance.

Multicollinearity was assessed to determine whether some independent variables carried overlapping information that could make coefficient estimates unstable. Collinearity in both the multiple and logistic regression models was evaluated using variance inflation factors (VIFs). Values of 5 or higher were considered indicative of problematic collinearity. In this study, no relevant multicollinearity was detected, as all VIF values were below 2.0. A summary of this analysis is provided in Multimedia Appendix 2.

Figure 1 shows a flowchart summarizing the analysis steps. As proof of concept, we conducted this preliminary analysis to assess the feasibility of the proposed approach. This analysis was not intended to provide definitive inference, but rather to demonstrate that the proposed method can be implemented and yield meaningful results under realistic conditions.

**Figure 1.** Flowchart summarizing the methodological workflow of the proof-of-concept study using data from the TSHA. EDA included histogram, box plot, and scatter plot visualizations. The FTS-12 score was analyzed using multiple linear regression, while the binary log-odds outcome was analyzed using logistic regression. A stepwise variable selection method was applied as an exploratory diagnostic tool, together with added model selection criteria. Bootstrapping procedures were incorporated into the multiple regression analysis, and ROC curve analysis was applied within the logistic regression framework. Multicollinearity assessment was conducted for both regression approaches. External model validation is pending and will be addressed in future work to further improve the variable selection process. EDA: exploratory data analysis; FTS-12: Frailty Trait Scale 12; RMS: residual mean square; ROC: receiver operating characteristic; TSHA: Toledo Study for Healthy Aging.






XSL•FO
RenderX

## Results

The results presented in this section correspond to an exploratory, proof-of-concept analysis and should be read accordingly: the estimates and discriminatory measures reported are intended to characterize the frailty-related signal contained in routine laboratory data, not to define validated clinical prediction tools. Their interpretation is therefore conditional on external validation in independent cohorts.

### Age and FTS-12 Values

The mean age of participants included in the selection process was 75.04 (SD 5.84; range 65 - 95) years for female participants and 74.95 (SD 5.63; range 65 - 99) years for male participants. The mean FTS-12 scores observed in the study were 46.68 (SD 15.99; range 6 - 92.11) for female participants and 39.62 (SD 15.43; range 0 - 84) for male participants. The original database comprised 847 (57%) female participants and 637 (43%) male participants.

### Models Based on FTS-12 Values: Selected Risk Factors and Functions

Tables S1 and S2 in Multimedia Appendix 3 summarize the variables retained in the regression models for female participants and male participants, respectively. In both sexes, the selected models combined individual laboratory markers with composite functions derived from clinically relevant variables. Table 1 presents the diagnostic statistics for the models stratified by sex. The *F*-statistics for both models were highly significant ($P<.001$), and the root mean square error ranged from 12 to 12.45. These findings highlight the relevance of both individual risk factors and their combined effects in modeling FTS-12 scores. Overall model performance was moderate, with $R^2$ values of approximately 0.39 in female participants and 0.38 in male participants. The parameter estimates and associated statistics for the selected multiple regression models in female and male participants are presented in Tables 2 and 3, respectively. The definitions and mathematical expressions of the composite risk factors used across the regression models are provided in Table 4.

Figures 2 and 3 further illustrate the modeling quality of the models. A moderately strong Pearson correlation was observed between the observed and modeled FTS-12 scores for both sexes, with correlations of 0.621 for female participants and 0.615 for male participants. The predicted FTS-12 scores were reasonably aligned with the observed scores in both sexes, supporting the feasibility of this approach while also indicating that substantial prediction uncertainty remains. The scatter plots visually represent this positive relationship, with a colored ellipsoid enclosing 90% of the data points. Higher correlations indicate better agreement between observed and predicted FTS-12 values.

**Table .** Descriptive characteristics and diagnostic analysis of selected multiple regression models for female and male participants[a].

| Statistics | Case study | |
|---|---|---|
| | Female | Male |
| *F* value | 46.81 (10, 744) | 37.22 (9, 550) |
| *P* value | <.001 | <.001 |
| Root MSE[b] | 12.45 | 12.00 |
| $R^2$ | 0.39 | 0.38 |
| Percentile 97.5% for $R^2$ | 0.436 | 0.493 |
| Percentile 2.5% for $R^2$ | 0.298 | 0.312 |
| Adjusted $R^2$ | 0.38 | 0.37 |
| Total of observations read | 847 | 637 |
| Total of observations used | 755 | 560 |
| Simple Pearson correlation between observed and expected values of FTS-12[c] | 0.621 | 0.615 |

[a]Percentiles 97.5% and 2.5% for $R^2$ were obtained by using the Bootstrapping technique. See the Modeling Regression Equations to Predict the FTS-12 Score section for further details.

[b]MSE: mean square error obtained from the ANOVA table of regression analyses.

[c]FTS-12: Frailty Trait Scale 12.



XSL•FO
RenderX

**Table .** Parameter estimates for the selected multiple regression model: female participants[a].

| Variable | Estimates | | | | | Correlations between FTS-12[b] versus risk factors in the model | | |
|---|---|---|---|---|---|---|---|---|
| | *df* | Estimate | SE | *t* test (*df*=744) | *P* value | Pearson correlation | *P* value | Values, n |
| Intercept | 1 | −337.16 | 32.57 | −10.35 | <.001 | —[c] | — | — |
| CRF1[d] | 1 | 25.40 | 2.16 | 11.74 | <.001 | 0.578 | <.001 | 825 |
| CRF2 | 1 | 19.67 | 3.96 | 4.97 | <.001 | 0.465 | <.001 | 839 |
| GGT[e] (U/L) | 1 | 0.1 | 0.037 | 3.18 | .002 | 0.120 | <.001 | 829 |
| Urea (mg/dL) | 1 | 0.15 | 0.048 | 3.17 | .002 | 0.178 | <.001 | 829 |
| Granulocytes (%) | 1 | 0.16 | 0.053 | 3.05 | .002 | 0.201 | <.001 | 838 |
| Creatinine (mg/dL) | 1 | −9.25 | 3.16 | −2.93 | .004 | 0.132 | <.001 | 840 |
| Bilirubin total (mg/dL) | 1 | −5.40 | 1.96 | −2.76 | .006 | −0.051 | .14 | 839 |
| Uric acid (mg/dL) | 1 | 1.11 | 0.42 | 2.62 | .009 | 0.187 | <.001 | 840 |
| Chlorine (mEq/L) | 1 | −0.30 | 0.16 | −1.92 | .05 | −0.126 | <.001 | 842 |
| Triglycerides (mg/dL) | 1 | 0.02 | 0.01 | 1.80 | .07 | 0.068 | .05 | 823 |
| Age (years) | — | — | — | — | — | 0.515 | <.001 | 847 |
| RDW[f] (%) | — | — | — | — | — | 0.240 | <.001 | 828 |
| Albumin (g/dL) | — | — | — | — | — | −0.389 | <.001 | 843 |

[a]CRF1 and CRF2 are composite risk factors defined in Table 4. Simple Pearson correlations between each risk factor in the selected reduced model and the FTS-12 score are also shown.

[b]FTS-12: Frailty Trait Scale 12.

[c]Not applicable.

[d]CRF: composite risk factor.

[e]GGT: gamma-glutamyl transferase.

[f]RDW: red cell distribution width.



XSL•FO
RenderX

**Table .** Parameter estimates for the selected multiple regression model: male participants[a].

| Variable | Estimates | | | | | Correlations between FTS-12[b] versus risk factors in the model | | |
|---|---|---|---|---|---|---|---|---|
| | *df* | Estimate | SE | *t* test (*df*=550) | *P* value | Pearson correlation | *P* value | Values, n |
| Intercept | 1 | 122.10 | 23.62 | 5.17 | <.001 | —[c] | — | — |
| CRF3[d] | 1 | −37.37 | 4.14 | −9.02 | <.001 | −0.552 | <.001 | 631 |
| CRF2 | 1 | 15.06 | 5.10 | 2.95 | .003 | 0.469 | <.001 | 625 |
| Chlorine (mEq/L) | 1 | −0.54 | 0.17 | −3.19 | .002 | −0.111 | .006 | 629 |
| GGT[e] (U/L) | 1 | −0.087 | 0.032 | −2.68 | .008 | −0.128 | .001 | 620 |
| Triglycerides (mg/dL) | 1 | 0.033 | 0.012 | 2.63 | .009 | 0.009 | .82 | 617 |
| Uric acid (mg/dL) | 1 | 0.94 | 0.39 | 2.41 | .02 | 0.080 | .046 | 625 |
| Phosphorus (mg/dL) | 1 | 2.48 | 1.04 | 2.37 | .02 | 0.048 | .23 | 626 |
| Eosinophil (%) | 1 | −0.56 | 0.29 | −1.90 | .05 | −0.040 | .32 | 623 |
| Leukocytes ($\times 10^3/\mu L$) | 1 | 0.60 | 0.33 | 1.86 | .06 | 0.090 | .02 | 626 |
| Age (years) | — | — | — | — | — | 0.476 | <.001 | 637 |
| Albumin (g/dL) | — | — | — | — | — | −0.427 | <.001 | 631 |
| Alkaline phosphatase (U/L) | — | — | — | — | — | 0.152 | .001 | 628 |

[a]CRF2 and CRF3 are composite risk factors defined in Table 4. Simple Pearson correlations between each risk factor in the selected reduced model and the FTS-12 score are also shown.

[b]FTS-12: Frailty Trait Scale 12.

[c]Not available.

[d]CRF: composite risk factor.

[e]GGT: gamma-glutamyl transferase.

**Table .** Definitions of composite risk factors (CRFs) used in the regression models[a].

| Label | Descriptor | Formula | Selected in model |
|---|---|---|---|
| CRF1 | RDW[b]-age-albumin composite | $\ln(RDW \times age^3/albumin)$ | Female participants, multiple regression |
| CRF2 | Alkaline phosphatase-age-albumin composite | $\ln(AlkPhos^c \times age/Albumin)$ | Female and male participants, multiple regression |
| CRF3 | Albumin-age composite | ln(albumin)×00/age | Male participants, multiple regression; female participants, logistic regression |
| CRF4 | Hemoglobin-age composite | $Hemoglobin \times 10^5/age^3$ | Male participants, logistic regression |

[a]Logarithmic transformations were applied to better capture nonlinear relationships between predictors and the Frailty Trait Scale 12 score, consistent with standard practice in regression modeling [21]. CRF identifiers are used throughout the paper; full mathematical definitions appear only in this table.

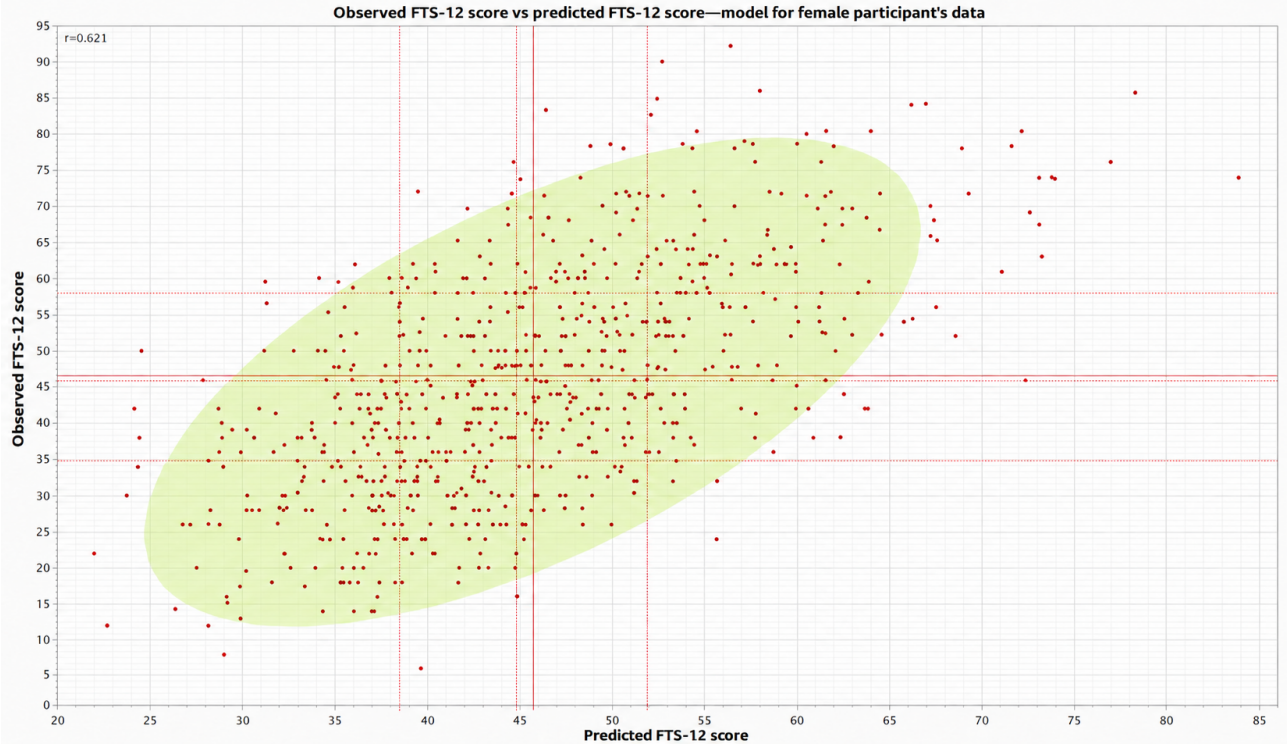
[b]RDW: red cell distribution width.

[c]AlkPhos: alkaline phosphatase.



XSL•FO
RenderX

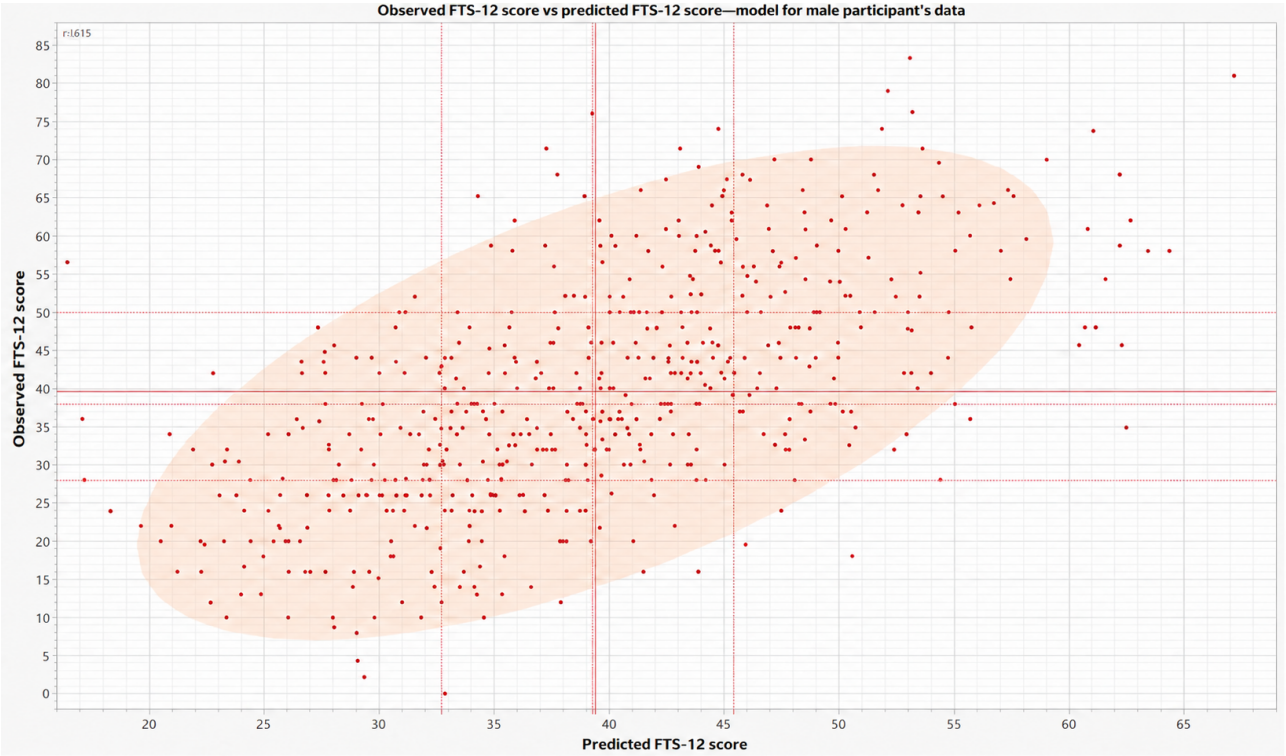
**Figure 2.** Observed versus modeled FTS-12 score using the selected multiple regression model in female participants (n=848). The horizontal axis shows the FTS-12 score predicted by the sex-specific multiple regression model (which retained age, albumin, gamma-glutamyl transferase, and selected composite risk factors as predictors; Table 2), and the vertical axis shows the FTS-12 score observed for each participant. The simple Pearson correlation was 0.621 ($P$<.001). Continuous vertical and horizontal red lines show the mean values for the modeled and observed FTS-12. Dashed vertical and horizontal red lines are the first (lower), second (median), and third (upper) quartiles. The shaded ellipse is the region encompassing 90% of the plotted data. A higher correlation, and closer alignment of points along the diagonal, indicate better agreement between observed and modeled FTS-12 values. The moderate dispersion observed is consistent with the multidimensional nature of frailty, only part of which is captured by routine laboratory data. FTS-12: Frailty Trait Scale 12.






XSL•FO
RenderX

**Figure 3.** Observed versus modeled FTS-12 scores using the selected multiple regression model in male participants (n=637). The horizontal axis shows the FTS-12 score predicted by the sex-specific multiple regression model (which retained age, albumin, gamma-glutamyl transferase, and a composite risk factor as predictors; Table 3), and the vertical axis shows the FTS-12 score observed for each participant. The simple Pearson correlation was 0.615 ($P$<.001). Continuous vertical and horizontal red lines show the mean values for the modeled and observed FTS-12. Dashed vertical and horizontal red lines are the first (lower), second (median), and third (upper) quartiles. The shaded ellipse is the region encompassing 90% of the plotted data. A higher correlation, and closer alignment of points along the diagonal, indicate better agreement between observed and modeled FTS-12 values. The moderate dispersion observed is consistent with the multidimensional nature of frailty, only part of which is captured by routine laboratory data. FTS-12: Frailty Trait Scale 12.




## Regression Analysis of Statistical Models

Tables 2 and 3 present the regression parameter estimates and related statistics for the 2 selected models. The composite risk functions were more strongly related to the FTS-12 score than most single laboratory variables. This suggests that frailty is better captured by combinations of biologically related markers than by isolated biomarkers alone. In female participants, the strongest association was observed for the function combining red cell distribution width (RDW), age, and albumin, whereas in male participants, the strongest association involved albumin- and age-related composite terms. These findings support the idea that frailty risk is influenced by interacting physiological processes rather than by a single laboratory abnormality. To enhance visualization, colored ellipsoids enclosing 90% of the data points are included in both figures. These ellipsoids highlight the trends that informed the selection of these functions for the models. Notably, the relationship is not strong enough to approximate a straight line. If they were, each value of the function composite risk factor (CRF) 1 (Figure S1 in Multimedia Appendix 4) or CRF3 (Figure S2 in Multimedia Appendix 4) would correspond to only 1 or a few closely related FTS-12 values. Instead, the ellipsoids suggest a broader range of FTS-12 scores for a specific value of each function. Nonetheless, a positive trend is evident for CRF1 (Figure S1 in Multimedia Appendix 4), while a negative trend is observed for CRF3 (Figure S2 in Multimedia Appendix 4). Comparatively, correlations with single risk factors were moderate. For female participants, the highest correlation was 0.187 ($P$<.001) for uric acid, with a positive correlation observed for gamma-glutamyl transferase (GGT; 0.120; $P$=.001). In contrast, GGT showed a negative correlation in male participants (−0.128; $P$=.001).

Correlations between FTS-12 and key variables such as age, RDW, albumin, and alkaline phosphatase were also examined. Age and albumin were of particular interest, as they were included in all CRF functions, regardless of sex. For age, the correlation was 0.515 ($P$<.001) in female participants and 0.476 ($P$<.001) in male participants. For albumin, the correlation was −0.389 ($P$<.0001) in female participants and −0.427 ($P$<.001) in male participants. Despite its moderately high correlation, albumin was not selected as an "individual variable." Plausible reasons for this exclusion are discussed later in the text.

## Models Based on High Frailty Probability: Selected Risk Factors and Functions

Tables S3 and S4 in Multimedia Appendix 3 summarize the statistics for the predicted probabilities (hereafter referred to as "the probabilities") and the independent variables, including individual risk factors and composite functions, incorporated into the selected logistic models for female participants and male participants, respectively. These tables also include information on "age" and several other risk factors. Although these variables were not selected as individual predictors, they are components of the CRF functions named through the selection process.



XSL•FO
RenderX

Table 5 provides descriptive statistics and diagnostic parameters for the selected logistic models. Among the participants, 200 of 735 (27%) in the female participants and 100 of 564 (18%) in the male participants had FTS-12 scores in the fourth quarter (above the third quartile, showing a high probability of frailty). The remaining participants in each group were distributed across the lower 3 quarters.

**Table .** Descriptive characteristics and diagnostic analysis of selected logistic regression models for female participants and male participants.

| Statistics | Case study | |
|---|---|---|
| | Female participants | Male participants |
| Observations read, n | 847 | 637 |
| Observations used, n | 735 | 564 |
| Observations with an FTS-12[a] value located in the fourth interval (larger than the third quartile), n | 200 | 100 |
| Observations with an FTS-12 value not located in the fourth interval (larger than the third quartile), n | 535 | 464 |
| Model fit statistics criterion | | |
| AIC[b] | 696.4 | 421.8 |
| −2 Log *L* | 682.4 | 403.8 |
| Max-rescaled $R^2$ | 0.312 | 0.323 |
| Testing global null hypothesis: only intercept | | |
| Likelihood ratio | | |
| Chi-square (*df*) | 178.0 (6) | 123.3 (10) |
| *P* value | <.001 | <.001 |

[a]FTS-12: Frailty Trait Scale 12.

[b]AIC: Akaike information criterion.

Tests of the global null hypothesis were highly significant (*P*<.001), showing that at least 1 independent variable—whether a single risk factor or a composite function of risk factors—was significantly associated with the probability of FTS-12 falling into the uppermost range.

Tables 6 and 7 summarize the estimated parameters and associated statistics for the selected models applied to female and male participants, respectively. The model for female participants includes 5 single variables and 1 composite function of risk factors. In contrast, the model for male participants comprises 7 single variables and 1 composite function of risk factors.



XSL•FO
RenderX

**Table .** Parameter estimates for the selected logistic model: female participants[a].

| Variable | Estimates | | | | | Correlations between probability versus risk factors in the model | | | Odds ratio estimates |
|---|---|---|---|---|---|---|---|---|---|
| | *df* | Estimate | SE | Wald chi-square (*df*=1) | *P* value | Pearson correlation | *P* value | Values, n | |
| Intercept | 1 | 7.83 | 1.43 | 30.0 | <.001 | __[b] | — | — | — |
| CRF3[c] | 1 | −6.21 | 0.63 | 96.9 | <.001 | −0.871 | <.001 | 735 | 0.00 |
| Granulocytes (%) | 1 | 0.035 | 0.011 | 10.2 | .001 | 0.349 | <.001 | 735 | 1.04 |
| Alkaline phosphatase (U/L) | 1 | 0.011 | 0.004 | 5.8 | .02 | 0.333 | <.001 | 735 | 1.01 |
| Urea (mg/dL) | 1 | 0.012 | 0.0082 | 2.1 | .14 | 0.248 | <.001 | 735 | 1.01 |
| LDL[d] (mg/dL) | 1 | −0.0057 | 0.0031 | 3.4 | .07 | −0.196 | <.001 | 735 | 0.99 |
| GGT[e] (U/L) | 1 | 0.021 | 0.0070 | 8.9 | .003 | 0.185 | <.001 | 735 | 1.02 |
| Age | — | — | — | — | — | 0.809 | <.001 | 735 | — |
| FTS-12[f] | — | — | — | — | — | 0.580 | <.001 | 735 | — |

[a]Simple Pearson correlations between each risk factor included in the selected model and the estimated probability of the FTS-12 values greater than the third quartile (fourth quarter) are also presented. CRF3 is a composite risk factor defined in Table 4. Pearson correlations between each risk factor and the modeled probability of falling above the FTS-12 third quartile are also presented, together with odds ratio estimates.

[b]Not available.

[c]CRF: composite risk factor (Table 4).

[d]LDL: low-density lipoprotein.

[e]GGT: gamma-glutamyl transferase.

[f]FTS-12: Frailty Trait Scale 12.



XSL•FO
RenderX

**Table .** Parameter estimates for the selected logistic model: male participants[a].

| Variable | Estimates | | | | | Correlations between probability versus risk factors in the model | | | Odds ratio estimates |
|---|---|---|---|---|---|---|---|---|---|
| | *df* | Estimate | SE | Wald chi-square (*df*=1) | *P* value | Pearson correlation | *P* value | Values, n | |
| Intercept | 1 | 11.16 | 5.27 | 4.5 | .03 | —[b] | — | — | — |
| CRF4[c] | 1 | −1.28 | 0.21 | 36.8 | <.001 | −0.641 | <.001 | 564 | 0.19 |
| Albumin | 1 | −2.70 | 0.53 | 26.2 | <.001 | −0.623 | <.001 | 564 | 0.07 |
| Alkaline phosphatase (U/L) | 1 | 0.013 | 0.0054 | 6.0 | .01 | 0.299 | <.001 | 564 | 1.00 |
| CHCM[d] (g/L) | 1 | −0.28 | 0.16 | 3.0 | .08 | −0.192 | <.001 | 564 | 0.76 |
| Hemoglobin (g/dL) | 1 | 0.46 | 0.11 | 16.4 | <.001 | −0.190 | <.001 | 564 | 1.58 |
| Triglycerides (mg/dL) | 1 | 0.012 | 0.0029 | 16.2 | <.001 | 0.182 | <.001 | 564 | 1.01 |
| Phosphorus (mg/dL) | 1 | 0.89 | 0.28 | 10.3 | .001 | 0.134 | <.001 | 564 | 2.43 |
| Lactate dehydrogenase (U/L) | 1 | 0.0017 | 0.0021 | 0.6 | .42 | 0.057 | .001 | 564 | 1.00 |
| Age | — | — | — | — | — | 0.707 | <.001 | 564 | — |
| FTS-12[e] | — | — | — | — | — | 0.538 | <.001 | 564 | — |

[a]Simple Pearson correlations between each risk factor included in the selected model and the estimated probability of the FTS-12 values greater than the third quartile of the distribution are also presented. CRF4 is a composite risk factor defined in Table 4. Pearson correlations between each risk factor and the modeled probability of falling above the FTS-12 third quartile are also presented, together with odds ratio estimates.

[b]Not applicable.

[c]CRF: composite risk factor (Table 4).

[d]CHCM: mean corpuscular hemoglobin concentration.

[e]FTS-12: Frailty Trait Scale 12.

Considering the Wald chi-square statistics, the presence of CRF3, granulocytes, and GGT is highly significant and essential in the model for female participants. For male participants, CRF4, albumin, hemoglobin, triglycerides, and phosphorus are identified as highly significant and critical associated factors. Tables 6 and 7 also present Pearson correlations between each selected independent single variable or composite function of risk factors and the modeled probability. Additionally, these tables include correlations between the modeled probability and both age and FTS-12 scores within the working datasets. In absolute terms, correlations involving composite functions of risk factors exceeded 0.64 (*P*<.001). Notably, the function CRF3 in female participants showed particularly strong correlations. Among the single risk factors, albumin showed a highly negative correlation in the selected model for male participants.

Figures S3 and S4 in Multimedia Appendix 4 illustrate the functional relationship between the modeled probability and the composite functions of risk factors with the highest correlations. However, these associations were nonlinear. For the logistic model applied to female participants, the relationship between the observed values of the function and the predicted probability followed an inverted logistic S-curve (Figure S3 in Multimedia Appendix 4), with a simple correlation of −0.871 (*P*<.001). When the value of the function exceeded 2.20, the probability fell below .20 (odds of 1 to 4). Conversely, when values of the function were less than 1.55, the probability exceeded .70 (odds of 7 to 3).

Similarly, for the logistic model in male participants, the relationship between function CRF4 and the modeled probability also followed an inverted logistic S-curve (Figure S4 in Multimedia Appendix 4), with a simple correlation of −0.641 (*P*<.001). Here, values of the function above 4.75 corresponded to probabilities lower than .20 (odds of 1 to 4), whereas values below 1.50 were associated with probabilities exceeding .70 (odds of 7 to 3). We also estimated the simple correlations between the modeled probability and both patient age and observed FTS-12 values in the original datasets. Age was consistently included in all composite functions of risk factors selected by the selection process. All correlations were highly significant (*P*<.001), irrespective of sex. Notably, correlations



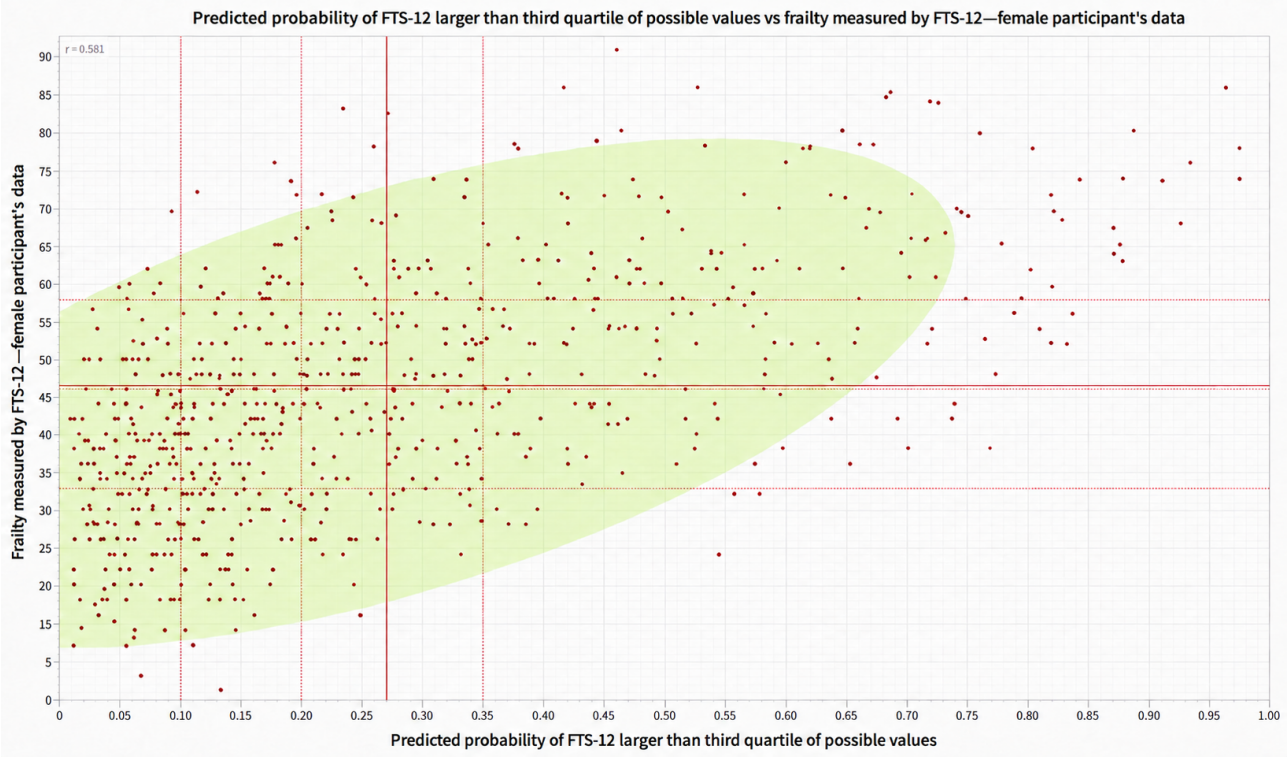
XSL•FO
RenderX

with age were stronger than those with FTS-12 (Tables 6 and 7).

Multimedia Appendix 4 also illustrates the functional relationship between the modeled probability and age in the selected models. For female participants, the scatter plot in Figure S5 in Multimedia Appendix 4 demonstrates that the distribution of values aligns closely with a logistic S-curve. Similarly, for male participants, Figure S6 in Multimedia Appendix 4 exhibits a positive trend, also resembling a logistic S-curve.

Positive correlations between the observed FTS-12 values and the modeled probabilities are clearer in the scatter plots presented in Figures 4 and 5 for male and female participants, respectively. While the dispersion of points in both plots shows a moderate correlation, a clear trend appears, highlighting the prediction capability of the selected models. This consistency in predictive performance is clear across both sexes.

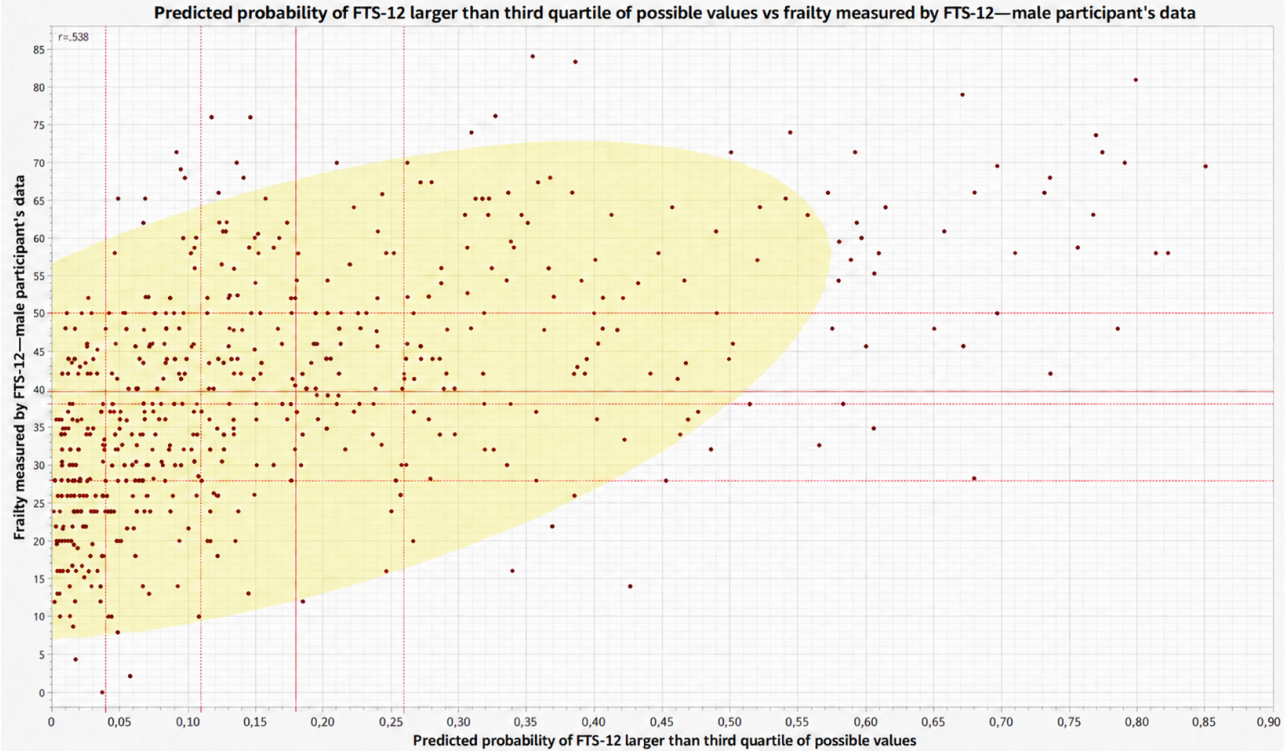
**Figure 4.** Scatter plot of observed FTS-12 values versus the modeled probability of the frailty scores larger than the third quartile of the distribution of possible FTS-12 values, based on the selected logistic model in female participants (n=848). The horizontal axis shows the modeled probability returned by the sex-specific logistic regression model (which retained albumin- and age-based composite risk factors and selected single biomarkers as predictors; Table 6), and the vertical axis shows the observed FTS-12 score for each participant. The probability quantifies the model-estimated likelihood that an individual falls within the highest-burden frailty subgroup of the cohort (ie, above the FTS-12 third quartile), used here as an internal analytical cutoff rather than as a clinical threshold. The Pearson correlation coefficient is 0.581 ($P$<.001). Continuous red lines are the mean values of the observed FTS-12 values and the modeled probability, while dashed red lines mark the first (lower), second (median), and third (upper) quartiles. The shaded ellipse is the region encompassing 90% of the plotted data. FTS-12: Frailty Trait Scale 12.






XSL•FO
RenderX

**Figure 5.** Scatter plot of observed FTS-12 values versus the modeled probability of the frailty scores larger than the third quartile of the distribution of possible FTS-12 values, based on the selected logistic model in male participants (n=637). The horizontal axis shows the modeled probability returned by the sex-specific logistic regression model (which retained hemoglobin- and age-based composite risk factors and selected single biomarkers as predictors; Table 7), and the vertical axis shows the observed FTS-12 score for each participant. The probability quantifies the model-estimated likelihood that an individual falls within the highest-burden frailty subgroup of the cohort (ie, above the FTS-12 third quartile), used here as an internal analytical cutoff rather than as a clinical threshold. The Pearson correlation coefficient is 0.538 ($P$<.001). Continuous red lines are the mean values of the observed FTS-12 values and the modeled probability, while dashed red lines mark the first (lower), second (median), and third (upper) quartiles. The shaded ellipse is the region encompassing 90% of the plotted data. FTS-12: Frailty Trait Scale 12.



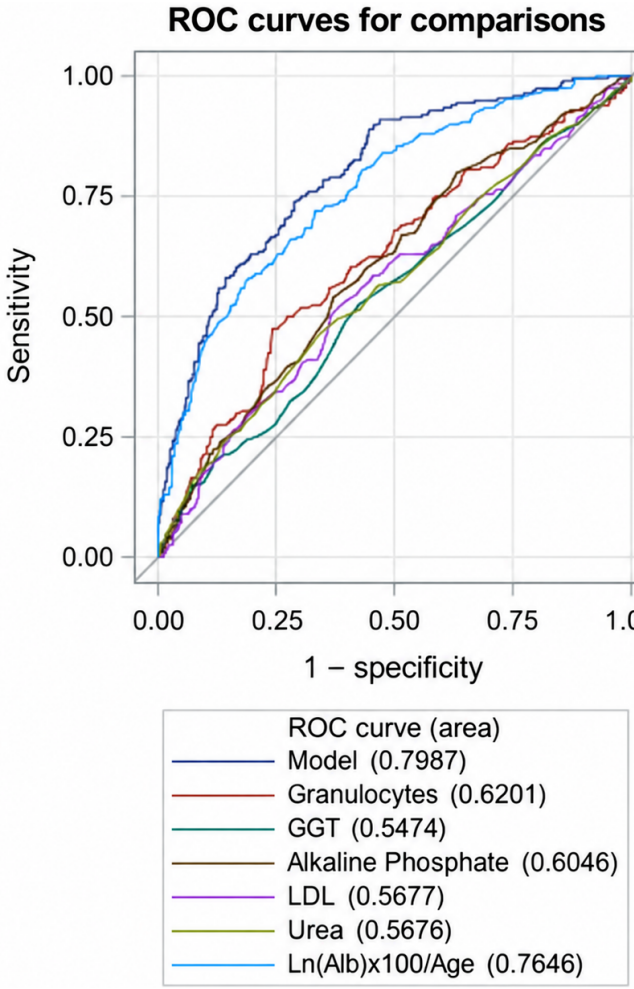

## Discriminatory Performance of the Logistic Models (ROC Analysis)

Using the AIC and −2Log *L* (minus 2 log-likelihood) to assess goodness-of-fit, the model for male participants proved a better fit compared to the model for female participants. Additionally, chi-square statistics derived from the likelihood ratio test provided strong evidence that the independent variables significantly influenced both selected models. The intercept alone was insufficient to explain the observed log-odds—the logarithm of the ratio between the probability of a FTS-12 value being larger than the third quartile of the distribution, and the probability of it not doing so. In addition to goodness-of-fit metrics, ROC curve analysis was used to evaluate the discriminatory performance of the selected logistic models. Tables S5 and S6 in Multimedia Appendix 3 provide statistics from the ROC analysis of the selected logistic models for both sexes. The ROC curves align closely with the goodness-of-fit statistics. For female participants, the AUC was approximately 0.799. Table S5 in Multimedia Appendix 3 and Figure 6 show that the function CRF3 was the most influential risk factor, accounting for 76% of the total AUC and closely approximating the model's overall AUC. For male participants, the AUC was approximately 0.824. Table S6 in Multimedia Appendix 3 and Figure 7 reveal that CRF4 and albumin were the most influential risk factors, contributing 73% and 71% of the total AUC, respectively.



XSL•FO
RenderX

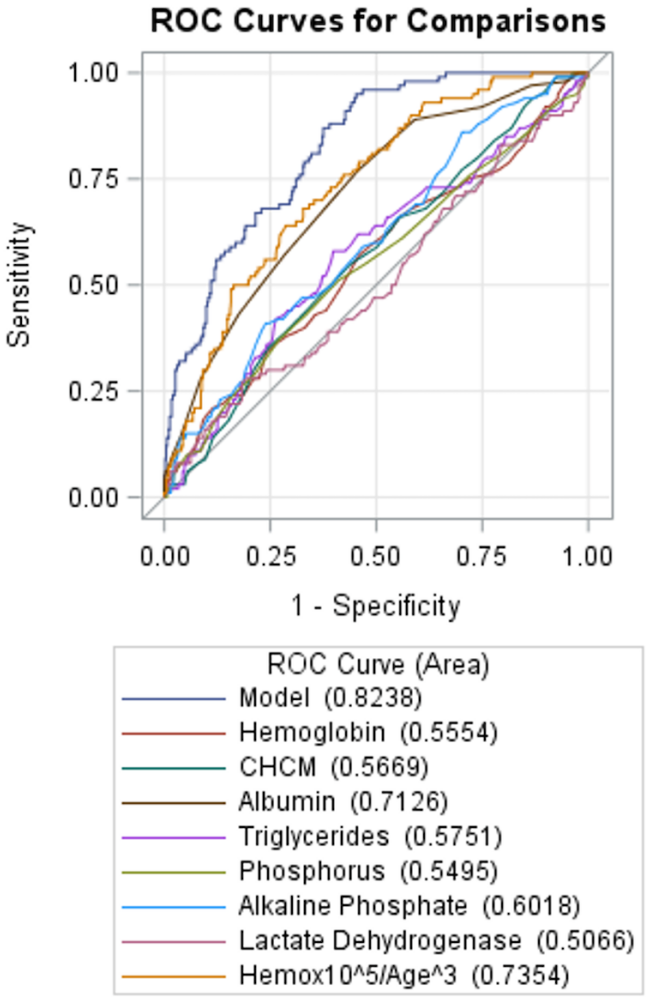
**Figure 6.** ROC curves for the selected logistic regression model and partial models in female participants (n=848). The ROC curve for the selected logistic regression model is shown alongside the ROC curves of partial models that include individual or composite risk factors for comparison. The curves illustrate the discriminatory power for the log-odds (and related probability) of an FTS-12 score in the uppermost range (above the third quartile) of the cohort. The horizontal axis shows 1–specificity (false positive rate), and the vertical axis shows sensitivity (true positive rate); the diagonal represents a noninformative classifier. The AUC for the selected model was 0.799, indicating moderate-to-good discrimination, and was substantially higher than the AUC of any model built on a single biomarker. The largest contribution among the partial models came from CRF3, accounting for approximately 76% of the total AUC. Key statistics, including the AUC for each model, highlight the improvement in predictive performance when combining risk factors in the selected model. CRF3 (the albumin-age composite) is defined in Table 4. AUC: area under the curve; CRF: composite risk factor; FTS-12: Frailty Trait Scale 12; GGT: gamma-glutamyl transferase; LDL: low-density lipoprotein; ROC: receiver operating characteristic.






XSL•FO
RenderX

**Figure 7.** ROC curves for the selected logistic regression model and partial models in male participants (n=637). The ROC curve for the selected logistic regression model is shown alongside the ROC curves of partial models that include individual or composite risk factors for comparison. The curves illustrate the discriminatory power for the log-odds (and related probability) of an FTS-12 score in the uppermost range (above the third quartile) of the cohort. The horizontal axis shows 1–specificity (false positive rate), and the vertical axis shows sensitivity (true positive rate); the diagonal represents a noninformative classifier. The AUC for the selected model was 0.824, indicating moderate-to-good discrimination, and was substantially higher than the AUC of any model built on a single biomarker. The largest individual contributions to the model AUC came from CRF4 (approximately 73% of the total AUC) and albumin (approximately 71%). Key statistics, including the AUC for each model, highlight the improvement in predictive performance when combining risk factors in the selected model. CRF4 (the hemoglobin-age composite) is defined in Table 4. AUC: area under the curve; CHCM: mean corpuscular hemoglobin concentration; CRF: composite risk factor; FTS-12: Frailty Trait Scale 12; ROC: receiver operating characteristic.




## Discussion

### Principal Findings

This exploratory study aimed to identify and refine sex-specific predictive equations for the FTS-12 score using routinely available biochemical and hematological data, including an analytical strategy to estimate the likelihood of belonging to the highest frailty range of the cohort. In line with these aims, the study showed that sex-specific multiple regression models could estimate the FTS-12 score with moderate accuracy, and that sex-specific logistic models could identify predictors associated with being in the upper range of the FTS-12 distribution. Across both approaches, age, albumin, GGT, and several composite variables emerged as recurrent contributors, suggesting that frailty-related laboratory signals are better captured through combinations of markers than through most single biomarkers alone. Taken together, these findings support the feasibility of a low-cost, laboratory-based approach for exploratory frailty risk stratification, but the present models do not constitute validated clinical prediction tools and would require external validation in independent cohorts before any clinical application can be considered.

### Interpretation of Findings and Comparison With Previous Literature

Our findings suggest that routinely available biochemical and hematological data can provide a meaningful, although still imperfect, estimate of frailty burden as measured by the FTS-12 score, as well as of the likelihood of being in the highest frailty range of the cohort. Some uncertainty in prediction is expected, as the FTS-12 reflects a multidimensional construct integrating 7 dimensions, including energy balance, nutrition, and vascular system components [11], rather than a single biological process. In this context, combinations of laboratory variables performed better than most isolated biomarkers, consistent with the stronger associations observed for the selected composite risk functions in both sexes (Tables 2 and 3 and Figures S3 and S4 in



XSL•FO
RenderX

Multimedia Appendix 4). The logarithmic transformations applied to variables such as albumin, age, and hemoglobin should be understood as a way to capture nonlinear clinical relationships and improve model fit, rather than as an end in themselves [21]. For example, combinations involving age and albumin may reflect the joint contribution of aging-related vulnerability and reduced nutritional or systemic reserve, in line with literature linking low albumin and hypoproteinemia to frailty, inflammation, and nutritional vulnerability in older adults [22,23]. Likewise, functions including RDW, hemoglobin, or albumin may capture overlapping signals related to erythrocyte heterogeneity, oxygen-carrying capacity, anemia-related vulnerability, and physiological resilience, consistent with evidence associating RDW and lower hemoglobin status with frailty in older populations [24,25]. In the logistic models, composite risk functions showed the strongest associations with higher frailty probability; and in male participants, albumin also emerged as an important contributor, supporting the interpretation of frailty as a multidimensional state of vulnerability shaped by interacting biological processes.

The models explained about 40% of the variability in FTS-12 scores, indicating moderate predictive performance. This level of explanation is clinically meaningful for an initial screening-oriented approach, while also showing that a substantial part of frailty remains influenced by factors not captured in laboratory-based models—a multidimensional syndrome reflecting biological, functional, cognitive, psychosocial, and environmental factors. The unexplained variance is therefore expected and reinforces that these models should be viewed as complementary tools rather than substitutes for comprehensive geriatric assessment, with future improvements depending on incorporating additional clinical, behavioral, and social determinants of frailty. Notably, albumin and age emerged as key associated factors consistently included across all composite functions, with logarithmic transformations capturing nonlinear relationships more effectively than models based solely on individual risk factors.

Recent literature includes only a few studies that model frailty risk using biochemical traits in older adults [21,23,26]. Several recent works have developed statistical models for frailty and frailty-associated outcomes such as mortality, most of them using logistic regression [27-30]. For example, Bencivenga et al [31] examined a FI composed of 38 multidimensional health deficits—including comorbidities, laboratory and diagnostic data, and symptoms and signs of disease—using procedures outlined by Searle et al [32]; the independent variables in their multiple regression model included sex (binary), age, BMI, and muscle thickness of the vastus intermedius and rectus femoris. Logistic regression is a widely used technique for modeling binary dependent variables in this setting [33-35]. Wang et al [36] provide a recent example using logistic regression and ROC curve analysis to compare 2 FIs in classifying binary mortality outcomes: the first based on routine physical assessments and laboratory tests, the second on self-reported data encompassing 30 health problems (“health deficits”) such as diseases, symptoms, disabilities, and psychological issues. In their study, FI values were used as independent variables to model the log-likelihood of dying. The clinical relevance of laboratory-based frailty stratification was further supported by Pasternak et al [37], who demonstrated that a laboratory-based FI was significantly associated with all-cause mortality in hospitalized older adults.

In contrast, our study used FTS-12 measurements to generate a new binary dependent variable, categorizing scores as exceeding the third quartile of the distribution or not. This allowed us to model the likelihood that an observed FTS-12 value would fall within the uppermost range, with logistic regression then applied to assess the log-likelihood of these outcomes. Liu et al [38](2021) also used logistic regression, but with a different approach: a binary outcome representing the presence or absence of frailty in older patients with nutritional risks, with frailty status determined from previous research. However, Liu et al [38] focused on exogenous (environmental) predictors such as age, sex, ethnicity, educational level, BMI, walking ability, activities of daily living, and cognitive functioning, whereas our study prioritized “endogenous” biochemical variables to identify both the “most reasonable” model and the “most influential” risk factors.

Álvarez-Sánchez et al [39] used logistic regression to evaluate the association between 2 inflammatory biomarkers—high-sensitivity C-reactive protein and homocysteine—and prefrailty in Spanish adults aged 65 years and older [3], underscoring the role of inflammation in age-related physical decline. Although neither biomarker was selected in our models, our findings on GGT and albumin are consistent with emerging geriatric evidence. In a Korean population-based cohort, Kang et al [26] reported that elevated GGT levels were 26% higher in frail male participants than in nonfrail male participants, with higher GGT quartiles showing a 2.08-fold increased odds of frailty. Elevated GGT has also been associated with poorer muscle function—including reduced jump power and diminished handgrip strength—independent of body-composition factors [40], and with increased risk of hip fracture in postmenopausal female participants, irrespective of alcohol consumption or chronic liver disease [41]. Because both diminished handgrip strength and elevated risk of hip fracture are strongly associated with frailty in older adults, these observations support the potential relevance of GGT in frailty risk estimation [3,11]. A recent systematic review further emphasized that hypoalbuminemia is associated with a significant risk of frailty and adverse health consequences in hospital and community contexts [24]. Together, these findings strengthen the biological plausibility of our results, suggesting that GGT and albumin reflect convergent pathways of oxidative stress, malnutrition, and chronic inflammation that are central to geriatric syndromes [42].

From a model-performance perspective, the diagnostic criteria used to compare candidate logistic models suggested acceptable exploratory fit [20,33,43]. More importantly, from a clinical standpoint, the models showed that higher frailty probability was better captured by combinations of laboratory variables than by most isolated biomarkers alone. This supports the idea that frailty-related vulnerability reflects interacting biological domains rather than a single abnormal test result and helps explain why composite terms contribute more consistently than most individual variables. Notably, the strongest correlations



XSL•FO
RenderX

were seen between CRF functions and both FTS-12 values and the associated probabilities, regardless of sex. However, substantial dispersion in observed FTS-12 values persisted for each value of the composite risk functions, highlighting the complex interplay of factors across the 7 dimensions underlying frailty assessment.

Placed in the context of existing approaches for laboratory-based frailty screening, the discriminatory performance reported in the Results section falls within the moderate-to-good range typically described for tools that rely exclusively on routine biochemical and hematological variables [11,21]. AUC values close to 0.80 for both sexes are consistent with what may reasonably be expected when functional, cognitive, and self-reported components are not incorporated, and they reinforce the view that the present models should be regarded as a complementary first layer of risk stratification rather than as a stand-alone diagnostic instrument. Beyond their overall discriminatory performance, the biological composition of the dominant composite functions is also informative: the prominence of the albumin-age composite in female participants is consistent with the well-established association of hypoalbuminemia and anemia with adverse outcomes in older adults [24,25], while the relevance of the hemoglobin-age composite in male participants is compatible, with erythropoietic decline acting as a sex-sensitive marker of physiological reserve. The recurrent involvement of GGT and alkaline phosphatase further suggests that subclinical hepatobiliary and oxidative pathways contribute to the laboratory signature of frailty [26], in line with the broader role of inflammatory, nutritional, and oxidative mechanisms previously implicated in the syndrome [3,39,42].

## Limitations

Variable selection was conducted using stepwise multiple regression guided by the AIC and Mallows $C_p$ statistic. Although stepwise procedures have recognized limitations, this approach was used for exploratory purposes within a proof-of-concept framework, and the results should therefore be interpreted with appropriate caution.

This study has several limitations that should temper the interpretation of its findings. First, its cross-sectional design prevents causal inference and may be susceptible to reverse causation; for example, elevated GGT could reflect unmeasured liver disease or alcohol intake rather than etiological pathways to frailty [26,44]. Second, all biochemical measurements were obtained at a single baseline visit, so temporal fluctuations and dynamic changes in frailty status were not evaluated. Third, the cohort was drawn from a single European population (Toledo, Spain), which may limit external validity in ethnically diverse or resource-limited settings. Finally, model development and internal validation were performed using the same dataset, which may increase the risk of overfitting despite the application of bootstrapping procedures. Future studies should therefore prioritize external or independent validation to assess model performance on unseen data and the generalizability of the proposed models.

Furthermore, rather than serving as a general clinical cutoff for the FTS-12 scale, the binary logistic analysis based on the upper quartile should be understood as an internal analytical technique to identify the subgroup with the highest frailty burden within this cohort. We recognize that dichotomization may reduce information and affect apparent model discrimination. The current proof-of-concept study did not include formal calibration analyses (calibration plots, calibration slope/intercept, and Brier score), threshold-specific sensitivity and specificity, decision-curve analysis, or net-benefit evaluation, although discrimination was investigated using ROC curve analysis and multicollinearity was confirmed to be nonproblematic using VIFs.

As noted in the Methods section, calibration in the probabilistic sense did not apply to the multiple regression model, for which performance was instead evaluated through goodness-of-fit measures and residual diagnostics. For the logistic models, formal calibration was not assessed. The present models should therefore be interpreted as exploratory, discrimination-based tools to determine whether routinely available laboratory data contain meaningful frailty-related signals, rather than as deployment-ready clinical prediction instruments. Further development toward clinical implementation will require calibration-focused evaluation, external validation, and implementation-oriented testing. Although bootstrap procedures were used to estimate CIs for selected model parameters, this does not constitute a full internal validation; optimism-corrected evaluation of overall model performance was beyond the scope of this secondary exploratory analysis.

More stringent development and validation techniques—penalized regression (eg, least absolute shrinkage and selection operator), repeated cross-validation, formal calibration evaluation, clinical utility analyses, and benchmarking against parsimonious and machine-learning alternatives—should be incorporated into model refinement in later phases. These procedures were outside the scope of the current work because the study was not preregistered and was intended as a secondary exploratory proof-of-concept analysis based on an existing cohort dataset. Complete-case analysis was used due to low missingness, and sex-stratified modeling was given precedence over formal interaction testing. Variables such as alcohol use, liver disease, medication exposure, inflammatory markers, nutritional variables, and comorbidity burden were not available in sufficiently complete form for inclusion, so part of the observed signal for biomarkers such as GGT and albumin may reflect unmeasured clinical contexts rather than direct frailty-specific pathways. The present revision does not yet provide TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis)-compliant reporting, optimism-corrected full model validation, or direct benchmarking against simpler baseline models (eg, age-only or age-plus-albumin), all of which should be incorporated in future validation-oriented refinement.

## Broader Implications

Our proof-of-concept analysis supports the viability of this modeling approach, and the next logical step is model refinement. A prospective, multicenter cohort with repeated biomarker assessments and adjudicated clinical outcomes would provide a more appropriate framework to confirm temporality



XSL•FO
RenderX

and improve generalizability. External validation in Latin American populations, integration of additional biological data layers, and comparison with machine-learning approaches may further improve predictive performance. Because the proposed models rely on routinely available laboratory parameters, they are particularly suitable for implementation in primary care and resource-limited settings, where they could help prioritize older adults for further geriatric assessment when specialized frailty evaluations are not immediately available. Embedding the frailty-probability equations within electronic health record dashboards or mobile screening tools could enable real-time, low-cost risk stratification in primary care and remote geriatric care. By facilitating earlier identification of risk patterns from routinely available laboratory data, this approach may contribute to more individualized and sex-sensitive frailty management strategies, consistent with the broader goal of precision geriatric care.

Recent discussions in clinical AI emphasize that real-world implementation requires explainability, fairness, human oversight, and careful workflow integration, so computational models should be developed to augment clinicians through earlier detection, smarter triage, and more targeted prevention rather than replace clinical judgment [45]. Likewise, work in adjacent computational health fields reinforces the need for interpretable pipelines, rigorous validation, and multicohort generalizability before routine implementation [46]. Taken together, these considerations position the present work not only as a statistical proof of concept but also as an early translational step toward clinically embedded, digitally supported frailty screening in aging care.

## Conclusions

From routinely available biochemical and hematological data, frailty, operationalized using the FTS-12 score, was modeled with moderate accuracy; sex-specific models demonstrated consistent predictive performance in both male and female participants. The final models included albumin, age, GGT, and a few composite risk functions that are biologically plausible correlates of frailty-related risk. For transparency and future validation, the final equations (albumin in g/dL and GGT in U/L) are given in Multimedia Appendix 1. Currently, these models should not be viewed as nomograms or screening tools with clinically useful biomarker thresholds, but rather as instruments for continuous frailty-risk estimation.

The models offer clinically interpretable signals that could support future workflow-integrated frailty risk stratification, despite the modest explained variance. This strategy may support digitally enabled screening, earlier frailty-oriented assessment, and prioritization for comprehensive geriatric assessment in primary care, geriatric care, and remote care settings following appropriate external and prospective validation. Rather than establishing definitive clinical decision rules, this work was designed to evaluate feasibility, computational stability, and interpretability as a secondary exploratory proof-of-concept analysis based on an existing cohort. Before regular clinical use, additional validation, calibration, and implementation-focused research will be necessary.

## Funding

This work was funded by Agencia Nacional de Investigación y Desarrollo (ANID)/FONDECYT (grant 3230783), the Interuniversity Center for Healthy Aging (Code RED211993), ANID/FONDECYT (grant 1260773), and ANID/FONDECYT (grant 1260612).

## Data Availability

The datasets analyzed during this study are not publicly available due to participant privacy protections and cohort-specific data-use agreements but may be available on reasonable request, subject to institutional review and approval by the Toledo Study for Healthy Aging data custodians. Requests for data access may be directed to the Research and Teaching Unit at Hospital Virgen del Valle in Toledo, Spain, with input from F José García García, principal investigator of the Toledo Study for Healthy Aging. No custom code was generated during this study.

## Authors' Contributions

Conceptualization: FZ, IP
Methodology: FZ, IP
Data curation: FZ, DA
Formal analysis: FZ
Investigation: FZ, DA, VV
Visualization: FZ
Project administration: FG, CA, EF, IP
Resources: FG, CA, EF, IP
Supervision: FG, CA, EF, IP
Validation: FG, CA, EF, IP
Writing—original draft: FZ, DA
Writing—review and editing: FZ, DA, VV



XSL•FO
RenderX

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Mathematical equations for the sex-specific multiple and logistic regression models.
[DOCX File, 73 KB - i-jmr_v15i1e76500_app1.docx ]

Multimedia Appendix 2
Statistical rationale for model development, variable selection, bootstrapping, and multicollinearity assessment.
[DOCX File, 605 KB - i-jmr_v15i1e76500_app2.docx ]

Multimedia Appendix 3
Descriptive statistics, selected risk factors, predicted probabilities, and receiver operating characteristic analysis results for the sex-specific regression models.
[DOCX File, 73 KB - i-jmr_v15i1e76500_app3.docx ]

Multimedia Appendix 4
Scatter plots illustrating relationships among composite risk factors, Frailty Trait Scale-12 scores, age, and modeled frailty probabilities in the sex-specific regression analyses.
[DOCX File, 30 KB - i-jmr_v15i1e76500_app4.docx ]

## References

1. Morley JE, Vellas B, van Kan GA, et al. Frailty consensus: a call to action. J Am Med Dir Assoc 2013 Jun;14(6):392-397. [doi: 10.1016/j.jamda.2013.03.022] [Medline: 23764209]
2. Martin FC, O'Halloran AM. Tools for assessing frailty in older people: general concepts. Adv Exp Med Biol 2020;1216(9-19):9-19. [doi: 10.1007/978-3-030-33330-0_2] [Medline: 31894542]
3. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001 Mar;56(3):M146-M156. [doi: 10.1093/gerona/56.3.m146] [Medline: 11253156]
4. Yamada M, Arai H. Understanding social frailty. Arch Gerontol Geriatr 2023 Dec;115:105123. [doi: 10.1016/j.archger.2023.105123] [Medline: 37473693]
5. Verghese J, Ayers E, Sathyan S, et al. Trajectories of frailty in aging: prospective cohort study. PLoS One 2021;16(7):e0253976. [doi: 10.1371/journal.pone.0253976] [Medline: 34252094]
6. Dibello V, Zupo R, Sardone R, et al. Oral frailty and its determinants in older age: a systematic review. Lancet Healthy Longev 2021 Aug;2(8):e507-e520. [doi: 10.1016/S2666-7568(21)00143-4] [Medline: 36098000]
7. Di Ciaula A, Portincasa P. The environment as a determinant of successful aging or frailty. Mech Ageing Dev 2020 Jun;188:111244. [doi: 10.1016/j.mad.2020.111244] [Medline: 32335099]
8. Hoogendijk EO, Afilalo J, Ensrud KE, Kowal P, Onder G, Fried LP. Frailty: implications for clinical practice and public health. Lancet 2019 Oct 12;394(10206):1365-1375. [doi: 10.1016/S0140-6736(19)31786-6] [Medline: 31609228]
9. Cesari M, Prince M, Thiyagarajan JA, et al. Frailty: an emerging public health priority. J Am Med Dir Assoc 2016 Mar 1;17(3):188-192. [doi: 10.1016/j.jamda.2015.12.016] [Medline: 26805753]
10. Rockwood K, Mitnitski A. Frailty in relation to the accumulation of deficits. J Gerontol A Biol Sci Med Sci 2007 Jul;62(7):722-727. [doi: 10.1093/gerona/62.7.722] [Medline: 17634318]
11. García-García FJ, Carcaillon L, Fernandez-Tresguerres J, et al. A new operational definition of frailty: the Frailty Trait Scale. J Am Med Dir Assoc 2014 May;15(5):371. [doi: 10.1016/j.jamda.2014.01.004] [Medline: 24598478]
12. Rodríguez-Mañas L, Féart C, Mann G, et al. Searching for an operational definition of frailty: a Delphi method based consensus statement: the frailty operative definition-consensus conference project. J Gerontol A Biol Sci Med Sci 2013 Jan;68(1):62-67. [doi: 10.1093/gerona/gls119] [Medline: 22511289]
13. García-García FJ, Carnicero JA, Losa-Reyna J, et al. Frailty Trait Scale-Short Form: a frailty instrument for clinical practice. J Am Med Dir Assoc 2020 Sep;21(9):1260-1266. [doi: 10.1016/j.jamda.2019.12.008] [Medline: 32005416]
14. Alonso-Bouzón C, Carcaillon L, García-García FJ, Amor-Andrés MS, El Assar M, Rodríguez-Mañas L. Association between endothelial dysfunction and frailty: the Toledo Study for Healthy Aging. Age (Dordr) 2014 Feb;36(1):495-505. [doi: 10.1007/s11357-013-9576-1] [Medline: 23959520]
15. Rosner B. Fundamentals of Biostatistics: Cengage Learning; 2015.
16. Forthofer RN, Lee ES, Hernandez M. Biostatistics: A Guide to Design, Analysis and Discovery: Academic Press; 2006.
17. Draper NR, Smith H. Applied Regression Analysis: Wiley; 1998.
18. Rawlings JO, Pantula SG, Dickey DA. Applied Regression Analysis: A Research Tool: Springer; 1998. [doi: 10.1007/b98890]
19. Searle SR. Linear Models: Wiley; 1997.



XSL•FO
RenderX

20. Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression: John Wiley & Sons; 2013.
21. Sepúlveda M, Arauna D, García F, Albala C, Palomo I, Fuentes E. Frailty in aging and the search for the optimal biomarker: a review. Biomedicines 2022 Jun 16;10(6):1426. [doi: 10.3390/biomedicines10061426] [Medline: 35740447]
22. Zhang L, Yang P, Yin F, Zhang J, Zhao B, Zhou J. Association between frailty and hypoproteinaemia in older patients: meta-analysis and systematic review. BMC Geriatr 2024 Aug 17;24(1):689. [doi: 10.1186/s12877-024-05275-9] [Medline: 39154175]
23. Riviati N, Indrajaya T, et al. Serum albumin as prognostic marker for older adults in hospital and community settings. Gerontol Geriatr Med 2024;10:23337214241249914. [doi: 10.1177/23337214241249914] [Medline: 38720941]
24. Palmer K, Vetrano DL, Marengoni A, et al. The relationship between anaemia and frailty: a systematic review and meta-analysis of observational studies. J Nutr Health Aging 2018;22(8):965-974. [doi: 10.1007/s12603-018-1049-x] [Medline: 30272101]
25. Li CM, Chao CT, Chen SI, Han DS, Huang KC. Elevated red cell distribution width is independently associated with a higher frailty risk among 2,932 community-dwelling older adults. Front Med (Lausanne) 2020;7:470. [doi: 10.3389/fmed.2020.00470] [Medline: 32984367]
26. Kang MG, Jung HW, Kim BJ. Elevated serum gamma-glutamyl transferase as a risk factor for frailty in older men: a nationwide population-based study. J Cachexia Sarcopenia Muscle 2025 Aug;16(4):e70034. [doi: 10.1002/jcsm.70034] [Medline: 40726312]
27. Gobbens RJJ, van der Ploeg T. The prediction of mortality by quality of life assessed with the WHOQOL-BREF: a longitudinal analysis at the domain and item levels using a seven-year follow-up period. Qual Life Res 2021 Jul;30(7):1951-1962. [doi: 10.1007/s11136-021-02790-9] [Medline: 33620614]
28. Gobbens RJJ, van Assen M, Augustijn H, Goumans M, van der Ploeg T. Prediction of mortality by the Tilburg Frailty Indicator (TFI). J Am Med Dir Assoc 2021 Mar;22(3):607. [doi: 10.1016/j.jamda.2020.07.033] [Medline: 32883597]
29. van der Ploeg T, Gobbens RJJ, Salem BE. Bayesian techniques in predicting frailty among community-dwelling older adults in the Netherlands. Arch Gerontol Geriatr 2023 Feb;105:104836. [doi: 10.1016/j.archger.2022.104836] [Medline: 36343439]
30. van der Ploeg T, Gobbens R. A comparison of different modeling techniques in predicting mortality with the Tilburg Frailty Indicator: longitudinal study. JMIR Med Inform 2022 Mar 30;10(3):e31480. [doi: 10.2196/31480] [Medline: 35353054]
31. Bencivenga L, Picaro F, Ferrante L, et al. Muscle ultrasound as imaging domain of frailty. Front Med (Lausanne) 2022;9:922345. [doi: 10.3389/fmed.2022.922345] [Medline: 35899217]
32. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. BMC Geriatr 2008 Sep 30;8(24):18826625. [doi: 10.1186/1471-2318-8-24] [Medline: 18826625]
33. Myers RH, Montgomery DC, Vining GG, Robinson TJ. Generalized Linear Models: With Applications in Engineering and the Sciences: Wiley; 2010.
34. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models: Springer; 2012.
35. Christensen R. Log-Linear Models and Logistic Regression: Springer; 2006.
36. Wang C, Fang X, Tang Z, et al. A frailty index based on routine laboratory data predicts increased risk of mortality in Chinese community-dwelling adults aged over 55 years: a five-year prospective study. BMC Geriatr 2022 Aug 17;22(1):679. [doi: 10.1186/s12877-022-03374-z] [Medline: 35978284]
37. Pasternak E, Freud T, Press Y. Association between the frailty index based on laboratory tests and all-cause mortality in hospitalized older adults: retrospective cohort study. JMIR Aging 2025 Sep 10;8:e70204. [doi: 10.2196/70204] [Medline: 40930060]
38. Liu H, Jiao J, Zhu M, et al. An early predictive model of frailty for older inpatients according to nutritional risk: protocol for a cohort study in China. BMC Geriatr 2021 Aug 18;21(1):465. [doi: 10.1186/s12877-021-02396-3] [Medline: 34407755]
39. Álvarez-Sánchez N, Álvarez-Ríos AI, Guerrero JM, et al. Homocysteine and C-reactive protein levels are associated with frailty in older Spaniards: the Toledo Study for Healthy Aging. J Gerontol A Biol Sci Med Sci 2020 Jul 13;75(8):1488-1494. [doi: 10.1093/gerona/glz168] [Medline: 31304964]
40. Lee S, Song D, Shin S, Hong N, Rhee Y. Elevated serum γ-glutamyl transferase is associated with low muscle function in adults independent of muscle mass. Nutrition 2022;103-104:111813. [doi: 10.1016/j.nut.2022.111813] [Medline: 36096057]
41. Kim KJ, Hong N, Yu MH, et al. Elevated gamma-glutamyl transpeptidase level is associated with an increased risk of hip fracture in postmenopausal women. Sci Rep 2022 Aug 17;12(1):13947. [doi: 10.1038/s41598-022-18453-9] [Medline: 35977988]
42. Arauna D, Navarrete S, Albala C, et al. Understanding the role of oxidative stress in platelet alterations and thrombosis risk among frail older adults. Biomedicines 2024 Sep 3;12(9):2004. [doi: 10.3390/biomedicines12092004] [Medline: 39335518]
43. Gbur EE, Stroup WW, McCarter KS, et al. Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences: Wiley; 2012.
44. Puukka K, Hietala J, Koivisto H, Anttila P, Bloigu R, Niemelä O. Age-related changes on serum ggt activity and the assessment of ethanol intake. Alcohol Alcohol 2006;41(5):522-527. [doi: 10.1093/alcalc/agl052] [Medline: 16855003]




XSL•FO
RenderX

45. Ogut E. Artificial intelligence in clinical medicine: challenges across diagnostic imaging, clinical decision support, surgery, pathology, and drug discovery. Clin Pract 2025 Sep 16;15(9):169. [doi: 10.3390/clinpract15090169] [Medline: 41002784]
46. Ogut E. Graph-theoretical mapping of cortical and subcortical network alterations in preclinical neurodegeneration. Discov Neurosci 2025;20(1):14. [doi: 10.1186/s13064-025-00214-9]

## Abbreviations

**AIC:** Akaike information criterion
**AUC:** area under the curve
**CRF:** composite risk factor
**FI:** Frailty Index
**FTS:** Frailty Trait Scale
**GGT:** gamma-glutamyl transferase
**RDW:** red cell distribution width
**ROC:** receiver operating characteristic
**TRIPOD:** Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
**TSHA:** Toledo Study for Healthy Aging
**VIF:** variance inflation factor

*Edited by M Balcarras; submitted 12.Aug.2025; peer-reviewed by A Sgaravatti, D Chavarro-Carvajal, E Ogut, I Barrientos-Calvo, L Meza, M Murgieri, T Tello; revised version received 29.May.2026; accepted 29.May.2026; published 30.Oct.2026.*

*Please cite as:*
*Zamudio F, Arauna D, Garcia F, Albala C, Villalobos V, Fuentes E, Palomo I*
*Modeling Frailty Risk Based on Biochemical and Hematological Parameters Using Multiple and Logistic Regression: Cross-Sectional Analysis of the Toledo Study for Healthy Aging*
*Interact J Med Res 2026;15:e76500*
*URL: https://www.i-jmr.org/2026/1/e76500*
*doi:10.2196/76500*

© Francisco Zamudio, Diego Arauna, Francisco Garcia, Cecilia Albala, Victoria Villalobos, Eduardo Fuentes, Ivan Palomo. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 30.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# COPD-SQ Positivity and Suspected Asthma Screen Positivity From a Representative, Community-Based Sample in Xishui County, Guizhou Province, China: Cross-Sectional Study

Yuhao Liu[1], BA; Shiyu Zhang[2], BSc; Zhoutao Zheng[1], BMed; Ke Huang[3,4,5], MD; Qiande Lai[6], BMed; Yingping Wang[7], BMed; Tianying Zhao[7], BM; Yingchi Luo[7], BMed; Xiangqin Lyu[6], BMed; Pascal Geldsetzer[8,9], PhD; Till Bärnighausen[1,10,11], MD, ScD; Ting Yang[3,4,5], MD; Simiao Chen[1,4,10*], ScD; Chen Wang[1,3,4,5*], MD

[1]School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijige Santiao 31, Dongcheng District, Beijing, China

[10]Faculty of Medicine and University Hospital, Heidelberg Institute of Global Health (HIGH), Heidelberg University, Heidelberg, Baden-Württemberg, Germany

[11]Department of Global and Population Health, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, United States

[2]School of Health Policy and Management, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[3]Department of Pulmonary and Critical Care Medicine, China-Japan Friendship Hospital, Beijing, China

[4]State Key Laboratory of Respiratory Health and Multimorbidity, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[5]National Center for Respiratory Medicine, Beijing, China

[6]Department of Pulmonary and Critical Care Medicine, People's Hospital of Xishui County, Zunyi, Guizhou, China

[7]Center for Disease Control and Prevention of Xishui County, Zunyi, Guizhou, China

[8]Division of Primary Care and Population Health, Department of Medicine, Stanford University, Stanford, CA, United States

[9]Chan Zuckerberg Biohub San Francisco, San Francisco, CA, United States

[*]these authors contributed equally

**Corresponding Author:**
Chen Wang, MD
School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijige Santiao 31, Dongcheng District, Beijing, China

## Abstract

**Background:** Chronic obstructive pulmonary disease (COPD) and asthma remain underdiagnosed in China, particularly in rural and resource-constrained settings. Questionnaire-based screening may support early risk identification; however, COPD-oriented and asthma-oriented screening approaches may capture different segments of the population. Single-disease screening may fail to identify the broader spectrum of people reporting respiratory symptoms or relevant exposures.

**Objective:** This study aimed to estimate the weighted proportions and the overlap between Chronic Obstructive Pulmonary Disease Screening Questionnaire (COPD-SQ) positivity and suspected asthma screen positivity in a community-based screening sample from Xishui County, China, and to characterize the demographic, exposure-related, and respiratory profiles of the resulting screening-defined groups.

**Methods:** We conducted a cross-sectional analysis of baseline screening data collected between June and December 2024 from permanent residents aged 35 years or older across all 26 townships in Xishui County, Guizhou Province, China. Permanent residents were defined as individuals who had lived in Xishui County for at least 3 months before the survey and were expected to remain in the same township within the following year. Screening questionnaires were completed online or with assistance from trained local staff during household visits. Analyses incorporated stratum-specific sampling weights proportional to population-to-sample ratios and accounted for township-level clustering, with linearized SEs. COPD risk was assessed using the COPD-SQ, with a score of 16 or higher indicating high risk. Suspected asthma screen positivity was defined as at least one positive response to a symptom item from the European Community Respiratory Health Survey (ECRHS). Participants were classified into 4 mutually exclusive groups: both screen-negative, COPD-SQ–positive only, suspected asthma screen-positive only, and dual screen-positive.

**Results:** Of 44,005 sampled residents, 41,839 completed the baseline screening, and 41,808 were included in the final analysis. The weighted proportions were 66.72% (95% CI 59.89% - 73.54%) for both screen-negative, 11.01% (95% CI 8.88% - 13.13%) for COPD-SQ–positive only, 10.97% (95% CI 8.92% - 13.02%) for suspected asthma screen-positive only, and 11.31% (95%



XSL•FO
RenderX

CI 8.27% - 14.34%) for dual screen-positive. Overall, 22.31% (95% CI 17.36% - 27.27%) were COPD-SQ–positive, 22.28% (95% CI 17.26% - 27.30%) were suspected asthma screen-positive, and 33.28% (95% CI 26.46% - 40.11%) screened positive on at least one instrument. The 2 screening approaches therefore identified similarly sized but only partly overlapping populations. The dual screen–positive group had the highest cumulative smoking exposure, biomass smoke exposure, prevalence of a family history of respiratory disease, and respiratory symptom burden, although several of these characteristics were components of the screening definitions.

**Conclusions:** Questionnaire-based COPD risk assessment and asthma symptom screening identified complementary populations in this community-based sample. These findings suggest that digitally enabled, integrated respiratory risk stratification may help organize demand for confirmatory assessment in resource-constrained primary care settings. However, a coherent care cascade pathway that follows screening should be in place to improve population health.

**International Registered Report Identifier (IRRID):** RR2-10.2196/85597

***(Interact J Med Res 2026;15:e108291)*** doi:10.2196/108291

**KEYWORDS**

chronic obstructive pulmonary disease; asthma; respiratory disease; community screening; risk assessment; primary health care; rural health; spirometry

## Introduction

Chronic respiratory diseases (CRDs) usually encompass chronic obstructive pulmonary disease (COPD), asthma, pneumoconiosis, interstitial lung disease, and pulmonary sarcoidosis. An estimated 468.3 million individuals were living with CRDs in 2021 [1]. COPD and asthma are the most common and harmful CRDs in China and are collectively referred to as obstructive lung diseases (OLDs) [2]. The prevalence of COPD was 13.7% among those aged 40 years and above [3], and that of asthma was 4.2% among those aged 20 years and above in the China Pulmonary Health Survey (CPHS) [4]. COPD also imposes an immense health economic burden on China, accounting for INT $1.363 trillion (constant 2017 international dollars, adjusted for purchasing power parity) from 2020 to 2050 [5]. One of the most significant challenges with COPD and asthma is that the prevalence of undiagnosed OLDs may be significantly higher than that of diagnosed cases—recent reviews have estimated that up to 70% of COPD and asthma cases are not diagnosed worldwide [6,7].

Awareness, early detection, and access to appropriate care for COPD and asthma remain limited among both general and high-risk populations in China. A recent study identified substantial unmet health care needs for COPD testing, formal diagnosis, treatment, and control in China, where only 41.0% had previously undergone lung function tests, 17.6% had formally been diagnosed with COPD, 8.5% had received treatment, and less than 5.0% had been controlled [8]. Similar gaps have been observed for asthma, where only 28.8% had been diagnosed with asthma by a physician, 23.4% had reported receiving a lung function test, and only 5.6% had been treated with inhaled corticosteroids [4].

These gaps are particularly pronounced in primary care and community settings, where health professionals may have limited experience, time, personnel, and diagnostic facilities for respiratory disease detection and management. Strengthening awareness and early detection of COPD and asthma at the primary care level is therefore essential. A previous modeling study has shown the cost-effectiveness of population-based COPD screening at the community level [9]. By using the Chronic Obstructive Pulmonary Disease Screening Questionnaire (COPD-SQ), we can identify individuals at high risk of COPD based on 7 items: age, cumulative smoking history, BMI, cough, breathlessness, family history of respiratory diseases, and exposure to biomass. The COPD-SQ has been validated among Chinese populations and has been recommended for COPD screening in primary care settings [10]. Asthma screening in primary care settings is rare because of the complex natural history and heterogeneity of presentations [11,12]. Multiple epidemiological studies have used standardized symptom-based screening questionnaires to identify patients with suspected asthma, and the screening tool used in the European Community Respiratory Health Survey (ECRHS) [13-15] for asthma has been validated in Chinese populations. The ECRHS asthma screening questionnaire has been widely used in prior studies, including the CPHS [4], Global Burden of Disease Study [16], and so on. Both the COPD-SQ and the ECRHS asthma screening questionnaire are feasible in primary care settings to identify individuals at high risk of COPD and patients with suspected asthma [4,9].

Population medicine seeks to maximize population health benefits and advance health equity by extending care beyond the treatment of established disease to prevention, early risk identification, diagnosis, and longitudinal management [17]. In resource-constrained primary care settings, however, screening organized around a single disease may identify only part of the population reporting respiratory symptoms or relevant exposures. An integrated approach that combines COPD-oriented risk assessment with asthma symptom screening may therefore be more consistent with a population medicine framework, particularly where access to quality-assured spirometry and specialist assessment is limited.

Digital health technologies offer an opportunity to implement such population-level respiratory risk stratification at scale. In the Population Medicine Multimorbidity Intervention in Xishui County (POPMIX) project [18-21], the COPD-SQ and ECRHS-derived asthma symptom items were delivered through a web-based questionnaire accessed by scanning a QR code. Residents could complete the questionnaire using their own



XSL•FO
RenderX

mobile devices, while trained local staff provided household-based assistance to those unable to complete it independently. Responses were captured electronically, enabling standardized information on respiratory symptoms and major exposures to be collected across geographically dispersed communities and to be potentially linked to referral, confirmatory assessment, and subsequent management. This study therefore goes beyond a direct comparison of 2 screening tools and examines whether their combined use can inform chronic respiratory risk stratification in a low-resource primary care setting. We aimed to estimate the weighted proportions and overlap of COPD-SQ positivity and suspected asthma screen positivity and to characterize the demographic, exposure-related, and respiratory profiles of the resulting screening-defined groups, with the broader goal of informing tiered allocation of confirmatory assessment.

## Methods

### Study Design and Setting

This cross-sectional analysis used screening data from the POPMIX project, a population-based community screening and management project conducted in Xishui County, Guizhou Province, southwestern China [18-21]. The POPMIX project was designed to identify and manage residents at high risk of COPD using a population medicine framework. This analysis focused on the screening stage and examined the joint distribution of COPD-SQ positivity and suspected asthma screen positivity.

This study was conducted from June 2024 to December 2024 in Xishui County, Guizhou Province, a mountainous county where access to health care services is usually difficult and where the smoking prevalence of Guizhou Province ranked second among all provinces in China [18-22]. Xishui County is characterized as a resource-limited county, where its gross domestic product per capita was 48% below the national average in 2024; however, it was designated as China's National Comprehensive Primary Care Experimental Area, which created opportunities for public health practice innovations.

### Study Participants

The initial screening of the POPMIX project used a government-provided roster of permanent residents aged 35 years or older across all 26 townships in Xishui County. Permanent residents were defined as individuals who had lived in Xishui County for at least 3 months and were expected to remain in the same township for the subsequent year. A total of 208,259 individuals met this definition and constituted the target population.

The sampling procedure was aligned with the cluster-randomized controlled trial design of the POPMIX project [18-21]. Townships were first classified as large or small according to whether their population size was above or below the township average and, combined with trial assignment, were grouped into 4 strata: 4 large intervention townships, 4 large control townships, 9 small intervention townships, and 9 small control townships. Residents were sampled within each stratum for screening. In this initial screening study, eligible participants were included if they (1) were adults aged 35 years or older; (2) were permanent residents in Xishui County; and (3) provided informed consent. Responses completed in 3 minutes or less, or with missing questionnaire items, were excluded from the analysis. For individuals unable to complete the electronic questionnaire independently, trained village doctors administered the questionnaire face to face, with interviews recorded for quality assurance. Stratum-specific sampling weights were applied to obtain estimates representative of permanent residents aged 35 years or older in Xishui County.

The sample size was determined by consensus between the Xishui County government and the academic team. To achieve broad coverage of permanent residents and ensure representativeness, 10,000 individuals were targeted within each stratum. Allowing for an anticipated attrition rate of 10%, 11,000 individuals were sampled per stratum, yielding a proposed total sample of 44,000. The minimum required sample size for the POPMIX trial was 2850 individuals at high risk of COPD and 3988 individuals with suspected asthma, as specified in the published protocols [18,20]. Given an estimated high-COPD-risk prevalence of approximately 25%, screening 44,000 individuals was expected to provide sufficient eligible participants. Thus, this sample size was considered adequate to meet both the POPMIX trial requirements and the local government's population coverage objective. Within each stratum, residents were selected using simple random sampling, with approximately 11,000 residents sampled per stratum based on the cluster randomized controlled trial sample-size calculation. In total, 44,005 residents aged 35 years or older were included in the baseline screening roster.

Of the 44,005 sampled residents, 2166 did not complete the baseline screening or did not provide usable questionnaire data (see *Results* section). Among the 41,839 respondents, 31 were excluded because of missing key items required to determine COPD-SQ or suspected asthma screening status, leaving 41,808 participants in the final analytic sample.

### Screening Instruments and Risk Group Definitions

In this study, we used QR codes to scale up the COPD-SQ and ECRHS asthma screening questionnaires in a low-resource setting in Xishui County, Guizhou Province, China. Participants could use their mobile phones to scan the QR codes and complete the screening questionnaires. This web-based digital screening can be easily scaled up and used to obtain data on risk factor exposures and symptoms in a cost-effective and efficient manner. Data monitoring procedures, including timing and recording, were applied to ensure the accuracy of the collected data. For residents in Xishui County who could not use mobile phones, village doctors administered face-to-face questionnaires to include them in the survey and ensure representation of the target population.

High-COPD-risk status was assessed using the COPD-SQ (ie, COPD-SQ–positive). The COPD-SQ is considered the most suitable questionnaire for the Chinese population. A previous study reported that, at a cutoff score of 16, the sensitivity, specificity, and correct classification rates for COPD diagnosis were 60.6%, 85.2%, and 82.7%, respectively [10]. Thus, a score of 16 or higher was classified as high-COPD-risk status (ie,



XSL•FO
RenderX

COPD-SQ–positive), whereas a score below 16 was classified as COPD-SQ negative. Suspected asthma screen positivity was assessed using symptom-based items derived from the ECRHS asthma screening questionnaire. The Chinese version of the ECRHS asthma screening questionnaire has been validated and used in several important nationwide and regional studies. The Chinese version of the ECRHS screening questionnaire has a sensitivity of 94.0% and a specificity of 71.0% if the respondent answers "yes" to any question item and a sensitivity of 86.0% and a specificity of 96.5% if the respondent answers "yes" to the question item of whether they have "wheezed in the past 12 months" [4,20]. Within our study, we defined patients with suspected asthma as participants who responded "yes" to any question and reported the prevalence of the "wheezing symptom" in Table S4 and Figures S1 and S2 in Multimedia Appendix 1.

Based on the combined COPD-SQ and ECRHS screening results, participants were classified into 4 mutually exclusive screening-defined risk groups: both screen-negative, COPD-SQ–positive only, suspected asthma screen-positive only, and dual screen-positive. The both screen–negative group was defined as COPD-SQ negative and ECRHS-negative; the COPD-SQ–positive only group as COPD-SQ–positive and ECRHS-negative; the suspected asthma screen–positive only group as COPD-SQ negative and ECRHS-positive; and the dual screen–positive group as COPD-SQ–positive and ECRHS-positive (see Table S1 in Multimedia Appendix 1). Dual screen positivity was defined solely by positive screening results on both instruments and should not be interpreted as a clinical diagnosis of asthma-COPD overlap.

## Ethical Considerations

This study forms an initial component of the POPMIX trials and was approved by the Ethics Committee of the Chinese Academy of Medical Sciences & Peking Union Medical College. The initial approvals were granted under approval numbers CAMS&PUMC-IEC-2024‐040 and CAMS&PUMC-IEC-2024‐041. Continuing ethical reviews were completed in June 2025 and June 2026, with updated approvals granted under CAMS&PUMC-IEC-2025‐061, CAMS&PUMC-IEC-2025‐062, CAMS&PUMC-IEC-2026‐080, and CAMS&PUMC-IEC-2026‐081. The study was conducted in accordance with the Declaration of Helsinki and subsequent amendments. Informed consent was obtained from all participants before enrollment. Participants identified as being at high risk of COPD or having symptoms suggestive of asthma were informed of their risk status through pop-up alerts following the initial screening and were advised to seek professional medical care. All data were stored at the Information Centre of Chinese Academy of Medical Sciences & Peking Union Medical College to ensure data safety. Data access is subject to approval by the study's principal investigator.

## Statistical Analysis

Initial sampling weights were calculated as the ratio of the target population size to the sampled population size within each of the 4 strata and were adjusted for inclusion in the analytic sample. Inclusion probabilities were estimated using weighted logistic regression with age group, sex, and township as categorical predictors. Final weights equaled the initial weight divided by the predicted inclusion probability; no trimming or additional calibration was applied.

All primary estimates used the final weights, with sample sizes reported as unweighted counts. Variance estimation accounted for stratification and township-level clustering. Screening proportions and 95% CIs were estimated using Taylor linearization and design-based *t* critical values. Agreement and Cohen κ were calculated from weighted joint and marginal proportions. Analyses using alternative definitions of suspected asthma used the same weights. Continuous variables were summarized using weighted means and SDs, and categorical variables were summarized using unweighted counts and weighted percentages. Participant characteristics were described overall and across the 4 screening-defined groups. Smoking duration, average packs smoked per day, and pack-years were calculated among ever-smokers only.

Stratum-specific sampling weights were calculated as the ratio of the target population size to the sampled population size within each of the 4 strata. Survey analyses accounted for the 4 sampling strata and township-level clustering, with SEs estimated using Taylor linearization. Weighted proportions and design-based 95% CIs were estimated for each screening-defined group and for the overall prevalence of COPD-SQ positivity, suspected asthma screen positivity, dual screen positivity, positivity on either screening approach, discordant screening results, and identical screening status.

The joint distribution of COPD-SQ and ECRHS-derived symptom screening results was examined using a weighted 2 × 2 cross-tabulation and visualized using a weighted mosaic plot. Weighted observed agreement and Cohen κ were reported descriptively to characterize the degree of overlap between the 2 screening approaches rather than to assess interchangeability.

To describe group-specific profiles, selected demographic, smoking-related, environmental exposure, family history, and respiratory symptom variables were summarized across the 4 screening-defined groups. For visualization, each characteristic was standardized across the 4 groups using *z* scores, while the original weighted means or percentages were retained in the heatmap labels. Because several characteristics were components of the COPD-SQ or ECRHS-derived screening definitions, these group differences were interpreted descriptively rather than as independent associations.

As part of a sensitivity analysis, suspected asthma screen positivity was redefined as positive responses to at least 2 ECRHS-derived symptom items, and the weighted proportions of suspected asthma screen positivity, dual screen positivity, and positivity on either screening approach were reestimated. All analyses were conducted using Stata (version 17.0; StataCorp LLC) and R software (version 4.1.2; R Foundation for Statistical Computing).

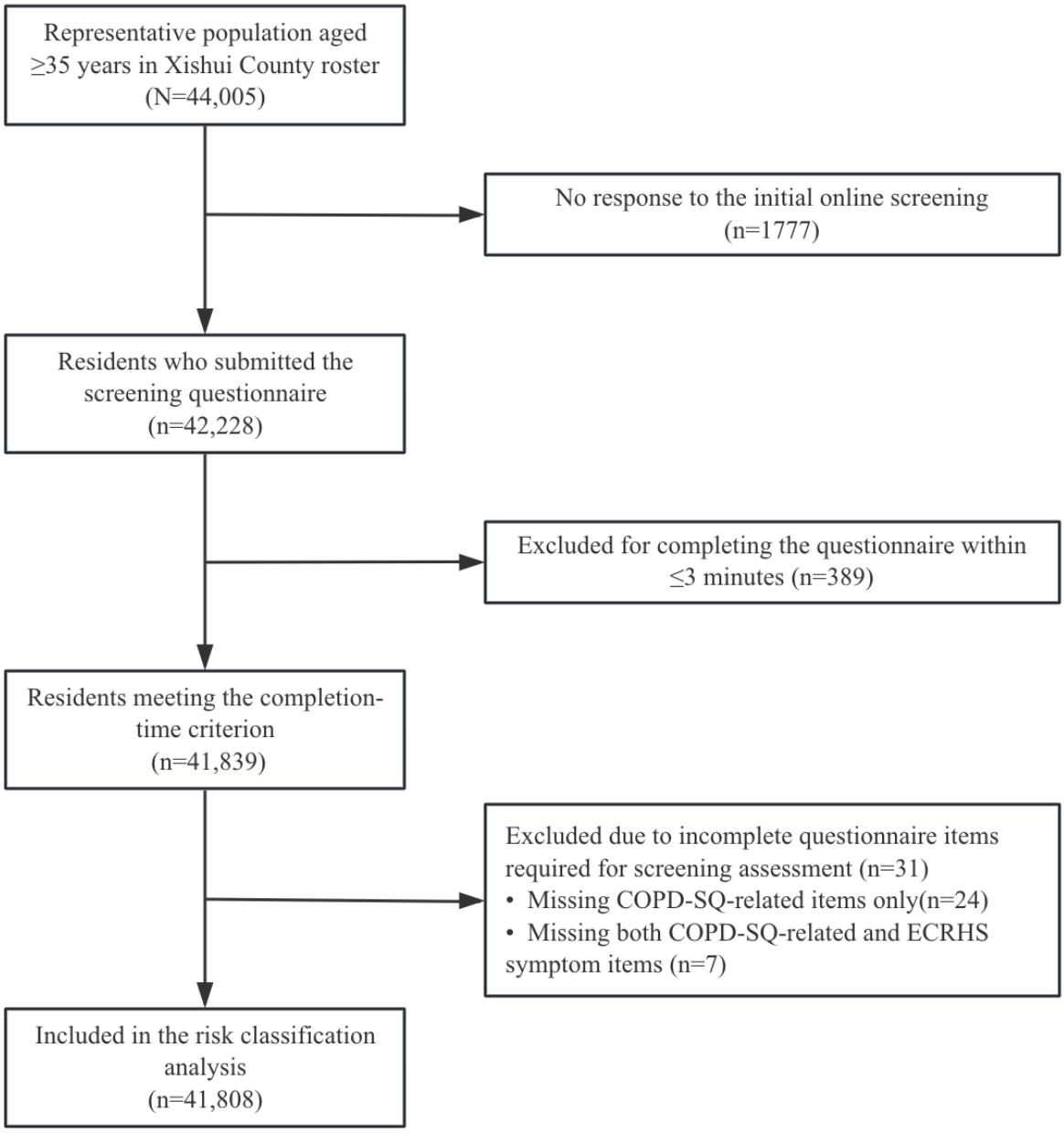


XSL•FO
RenderX

## Results

Among 44,005 permanent residents aged 35 years and older identified from a government-provided community roster in Xishui County, 2166 were excluded because of nonresponse, questionnaire completion time ≤3 minutes, or insufficient information to determine the screening group. We further excluded 31 respondents with missing key COPD-SQ or ECRHS items required for risk classification. Therefore, 41,808 participants were included in the final analytic sample (Figure 1). Of these, 27,542 participants were classified as both screen-negative, 4775 as COPD-SQ–positive only, 4628 as suspected asthma screen-positive only, and 4863 as dual screen-positive. Dual screen positivity was defined by positive results on both screening approaches and should not be interpreted as a clinical diagnosis of asthma-COPD overlap.

**Figure 1.** Flowchart of participant inclusion and exclusion. Participants were identified from a government-provided roster of permanent residents aged 35 years or older in Xishui County. Individuals who did not respond to the initial online screening had no screening data. Respondents missing key Chronic Obstructive Pulmonary Disease Screening Questionnaire (COPD-SQ) or European Community Respiratory Health Survey (ECRHS) items required for risk classification were excluded from the risk classification analysis. Permanent residents were defined as individuals who had resided in Xishui County for at least 3 months and were expected to remain within the same township for the subsequent year.




The 4 screening-defined groups showed distinct demographic, exposure-related, and respiratory symptom profiles (Table 1). Overall, the weighted mean age of participants was 57.75 (SD 12.86) years, and 50.79% (weighted) and 50.82% (21,247/41,808, unweighted) of participants were male. The COPD-SQ–positive only group was the oldest, with a weighted mean age of 71.11 (SD 8.73) years, and had the highest proportion of males (69.97% weighted and 70.01% unweighted [3343/4775]). By contrast, the suspected asthma screen-positive–only group had a weighted mean age of 54.55



XSL•FO
RenderX

(SD 10.69) years, the lowest proportion of males (41.06% weighted and 40.90% unweighted [1893/4628]), and the highest weighted mean BMI of 24.07 (SD 3.96) kg/m$^2$.



XSL•FO
RenderX

**Table .** Weighted characteristics of study population by screening-defined risk group.[a]

| Characteristics | Overall (N=41,808) | Both screen-negative (n=27,542) | COPD-SQ[b] positive only (n=4775) | Suspected asthma screen-positive only (n=4628) | Dual screen-positive (n=4863) |
|---|---|---|---|---|---|
| Age (y), weighted mean (SD) | 57.75 (12.86) | 54.19 (11.60) | 71.11 (8.73) | 54.55 (10.69) | 68.86 (10.16) |
| Height (m), weighted mean (SD) | 1.61 (0.07) | 1.62 (0.07) | 1.62 (0.07) | 1.60 (0.07) | 1.61 (0.07) |
| Weight (kg), weighted mean (SD) | 59.80 (10.39) | 60.63 (9.87) | 56.08 (9.97) | 61.79 (11.36) | 56.98 (11.35) |
| BMI ($kg/m^2$), weighted mean (SD) | 22.90 (3.49) | 23.17 (3.24) | 21.20 (3.13) | 24.07 (3.96) | 21.93 (3.89) |
| Cumulative years of smoking, weighted mean (SD) | 24.85 (13.71) | 19.01 (10.23) | 33.16 (13.95) | 20.88 (10.34) | 34.14 (13.75) |
| Average packs smoked per day, weighted mean (SD) | 0.95 (0.70) | 0.91 (0.66) | 0.96 (0.73) | 0.99 (0.64) | 1.05 (0.77) |
| Smoking index, pack-years, weighted mean (SD) | 23.91 (22.88) | 17.63 (16.48) | 31.25 (26.80) | 20.83 (16.69) | 35.15 (29.10) |
| Sex, n (weighted %) | | | | | |
| Male | 21,247 (50.79) | 13,128 (47.77) | 3343 (69.97) | 1893 (41.06) | 2883 (59.36) |
| Female | 20,561 (49.21) | 14,414 (52.23) | 1432 (30.03) | 2735 (58.94) | 1980 (40.64) |
| Smoking status, n (weighted %) | | | | | |
| Never smoked | 27,966 (66.99) | 20,518 (74.35) | 1762 (36.81) | 3340 (71.99) | 2346 (48.14) |
| Former smoker | 4822 (11.48) | 2247 (8.21) | 1096 (22.95) | 486 (10.64) | 993 (20.45) |
| Current smoker | 9020 (21.52) | 4777 (17.45) | 1917 (40.24) | 802 (17.37) | 1524 (31.41) |
| Cough without cold, n (weighted %) | | | | | |
| No | 37,530 (90.02) | 27,119 (98.49) | 4066 (85.16) | 4073 (88.03) | 2272 (46.67) |
| Yes | 4278 (9.98) | 423 (1.51) | 709 (14.84) | 555 (11.97) | 2591 (53.33) |
| Degree of dyspnea, n (weighted %) | | | | | |
| No dyspnea | 32,640 (78.56) | 25,507 (92.81) | 3148 (65.97) | 2792 (60.41) | 1193 (24.34) |
| Short of breath when walking normally on level ground | 2370 (5.51) | 388 (1.35) | 513 (10.64) | 242 (5.18) | 1227 (25.32) |
| Short of breath when hurrying on the level or walking up a slight hill | 6798 (15.93) | 1647 (5.83) | 1114 (23.40) | 1594 (34.41) | 2443 (50.35) |
| Exposure to biomass smoke from cooking or heating, n (weighted %) | | | | | |
| No | 34,712 (83.86) | 24,685 (90.28) | 3503 (73.69) | 3480 (76.15) | 3044 (63.40) |
| Yes | 7096 (16.14) | 2857 (9.72) | 1272 (26.31) | 1148 (23.85) | 1819 (36.60) |
| Family history of respiratory disease, n (weighted %) | | | | | |
| No | 39,340 (94.12) | 27,015 (98.03) | 4513 (94.43) | 4095 (88.31) | 3717 (76.42) |
| Yes | 2468 (5.88) | 527 (1.97) | 262 (5.57) | 533 (11.69) | 1146 (23.58) |
| Experienced single or recurrent wheezing, n (weighted %) | | | | | |
| No | 37,850 (90.72) | 27,542 (100) | 4775 (100) | 3180 (68.77) | 2353 (48.23) |
| Yes | 3958 (9.28) | 0 (0) | 0 (0) | 1448 (31.23) | 2510 (51.77) |
| Often bothered by cough at night, n (weighted %) | | | | | |



XSL•FO
RenderX

| Characteristics | Overall (N=41,808) | Both screen-negative (n=27,542) | COPD-SQ[b] positive only (n=4775) | Suspected asthma screen-positive only (n=4628) | Dual screen-positive (n=4863) |
|---|---|---|---|---|---|
| No | 38,759 (92.88) | 27,542 (100) | 4775 (100) | 3783 (81.89) | 2659 (54.59) |
| Yes | 3049 (7.12) | 0 (0) | 0 (0) | 845 (18.11) | 2204 (45.41) |
| Wheezing, dyspnea, or cough after exercise, n (weighted %) | | | | | |
| No | 36,073 (86.59) | 27,542 (100) | 4775 (100) | 2285 (49.70) | 1471 (30.23) |
| Yes | 5735 (13.41) | 0 (0) | 0 (0) | 2343 (50.30) | 3392 (69.77) |
| Symptoms after exposure to allergens or polluted air, n (weighted %) | | | | | |
| No | 38,311 (91.84) | 27,542 (100) | 4775 (100) | 3075 (66.64) | 2919 (60.23) |
| Yes | 3497 (8.16) | 0 (0) | 0 (0) | 1553 (33.36) | 1944 (39.77) |
| Cold symptoms go to the chest or last more than 10 days, n (weighted %) | | | | | |
| No | 37,339 (89.53) | 27,542 (100) | 4775 (100) | 2613 (56.50) | 2409 (49.57) |
| Yes | 4469 (10.47) | 0 (0) | 0 (0) | 2015 (43.50) | 2454 (50.43) |

[a]Values are n (weighted %) or weighted mean (SD). Counts are unweighted; percentages and continuous summaries are weighted using sampling weights adjusted for final analytic inclusion using roster age group, sex, and township. Smoking duration, packs/day, and pack-years were calculated among ever-smokers only. Heights outside 1.30 - 2.05 m, weights outside 25 - 250 kg, suspected weight-unit errors, and BMI values outside 10 - 60 kg/m$^2$ after cleaning were set to missing. COPD-SQ and ECRHS symptom rows are screening items and partly define the risk groups.

[b]COPD-SQ: Chronic Obstructive Pulmonary Disease Screening Questionnaire.

Smoking-related exposures varied across the screening-defined groups. Among ever-smokers, the dual screen–positive group had the highest mean smoking index at 35.15 (SD 29.10) pack-years, compared with 31.25 (SD 26.80) pack-years in the COPD-SQ–positive only group and 20.83 (SD 16.69) pack-years in the suspected asthma screen–positive only group. Current smoking was most common in the COPD-SQ–positive only group (1917/4775, 40.15%, unweighted and 40.24%, weighted), followed by the dual screen–positive group (1524/4863, 31.34%, unweighted and 31.41%, weighted). Exposure to biomass smoke from cooking or heating was most frequent in the dual screen–positive group (1819/4863, 37.40%, unweighted and 36.60%, weighted), followed by the COPD-SQ–positive only group (1272/4775, 26.64% unweighted and 26.31% weighted) and the suspected asthma screen–positive only group (1148/4628, 24.81% unweighted and 23.85% weighted).

A family history of respiratory disease was reported by 23.58% (weighted) and 23.57% (1146/4863, unweighted) of the dual screen–positive group, compared with 5.57% (weighted) and 5.49% (262/4775, unweighted) of the COPD-SQ–positive only group and 11.69% (weighted) and 11.52% (533/4628, unweighted) of the suspected asthma screen–positive only group. The dual screen–positive group also had the highest frequencies of cough, dyspnea, wheeze, nocturnal cough, exercise-related respiratory symptoms, allergen- or polluted air–triggered symptoms, and prolonged chest symptoms after colds. Because age, BMI, smoking exposure, cough, dyspnea, biomass smoke exposure, family history, and ECRHS-derived symptoms were components of the screening instruments, these patterns should be interpreted as descriptive characteristics of the screening-defined groups rather than as independent risk-factor associations.

The COPD-SQ and ECRHS-derived symptom screening identified partially overlapping but nonidentical populations (Figure 2). Among the 41,808 participants included in the analysis, 4775 were COPD-SQ–positive only, 4628 were suspected asthma screen-positive only, and 4863 were dual screen-positive. Overall, 22.31% (95% CI 17.36% - 27.27%) were COPD-SQ–positive and 22.28% (95% CI 17.26% - 27.30%) were suspected asthma screen-positive (Table 2). Overall, 33.28% (95% CI 26.46% - 40.11%) were screened positive on at least one of the 2 screening approaches (Table 2). Cross-classification showed that the corresponding weighted proportions were 11.01% (95% CI 8.88% - 13.13%), 10.97% (95% CI 8.92% - 13.02%), and 11.31% (95% CI 8.27% - 14.34%), respectively. The weighted observed agreement between the 2 screening approaches was 78.03% (95% CI 74.18%-81.87%), with a Cohen κ of 0.366 (see Table S2 in Multimedia Appendix 1). These findings indicate that the 2 approaches identified similarly sized but only partly overlapping populations, with substantial numbers identified by only one approach.



XSL•FO
RenderX

**Figure 2.** Weighted mosaic plot of Chronic Obstructive Pulmonary Disease Screening Questionnaire (COPD-SQ) and European Community Respiratory Health Survey (ECRHS) symptom screening. Rectangles' areas are proportional to weighted percentages based on sampling weights adjusted for final analytic inclusion. Both screen-negative denotes participants who are negative on both screening tools, and dual screen-positive denotes participants who are positive on both COPD-SQ and ECRHS symptom screening. Dual screen-positive should not be interpreted as physician-diagnosed asthma-COPD overlap.

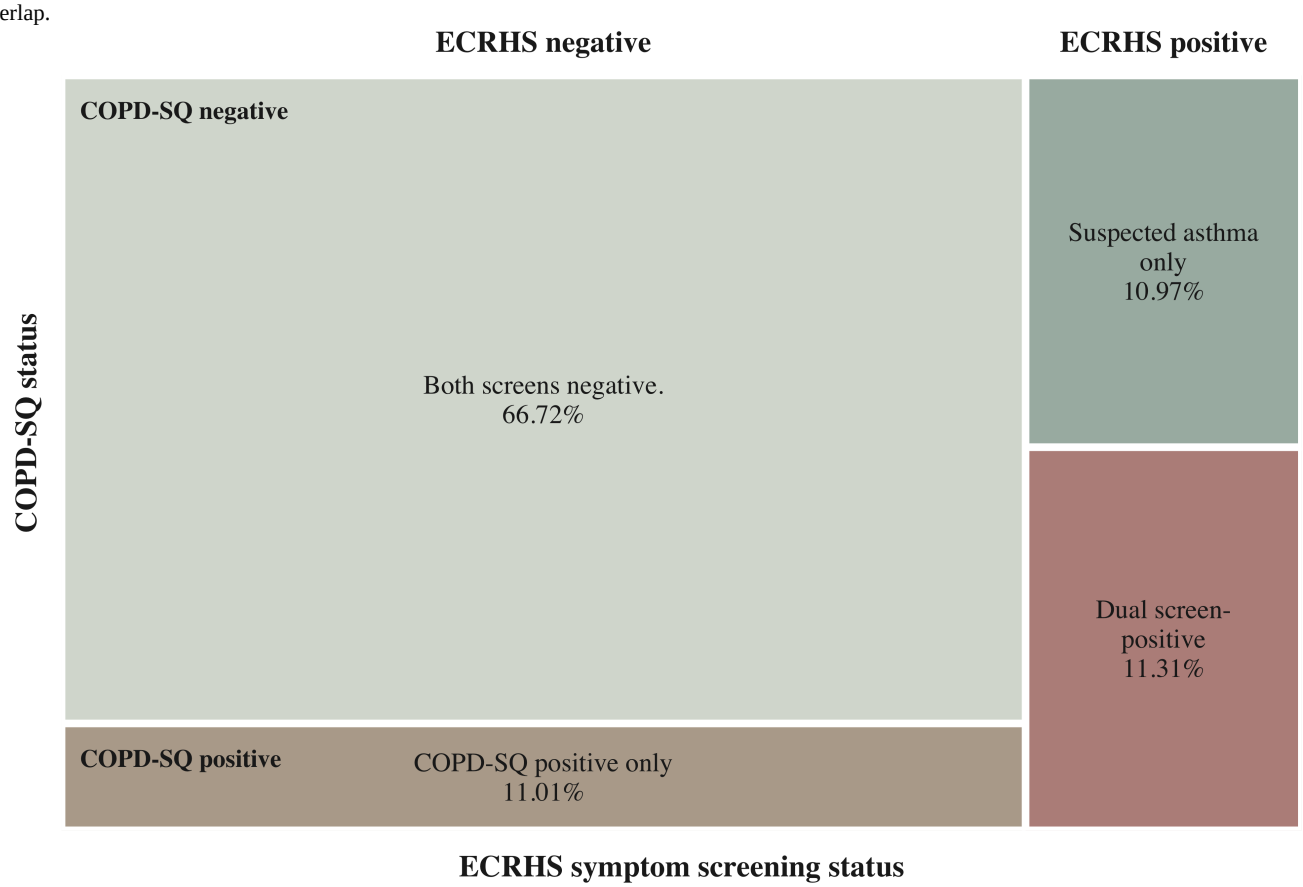



**Table .** Weighted prevalence estimates of Chronic Obstructive Pulmonary Disease Screening Questionnaire (COPD-SQ) and European Community Respiratory Health Survey (ECRHS) screening status.[a]

| Estimate | Unweighted n (weighted %; 95% CI [%]) |
|---|---|
| COPD-SQ positive | 9638 (22.31; 17.36 - 27.27) |
| Suspected asthma screen-positive | 9491 (22.28; 17.26 - 27.30) |
| Dual screen-positive | 4863 (11.31; 8.27 - 14.34) |
| Either COPD-SQ or ECRHS screen-positive | 14,266 (33.28; 26.46 - 40.11) |
| Discordant screen-positive | 9403 (21.97; 18.13 - 25.82) |
| Same screening status | 32,405 (78.03; 74.18 - 81.87) |

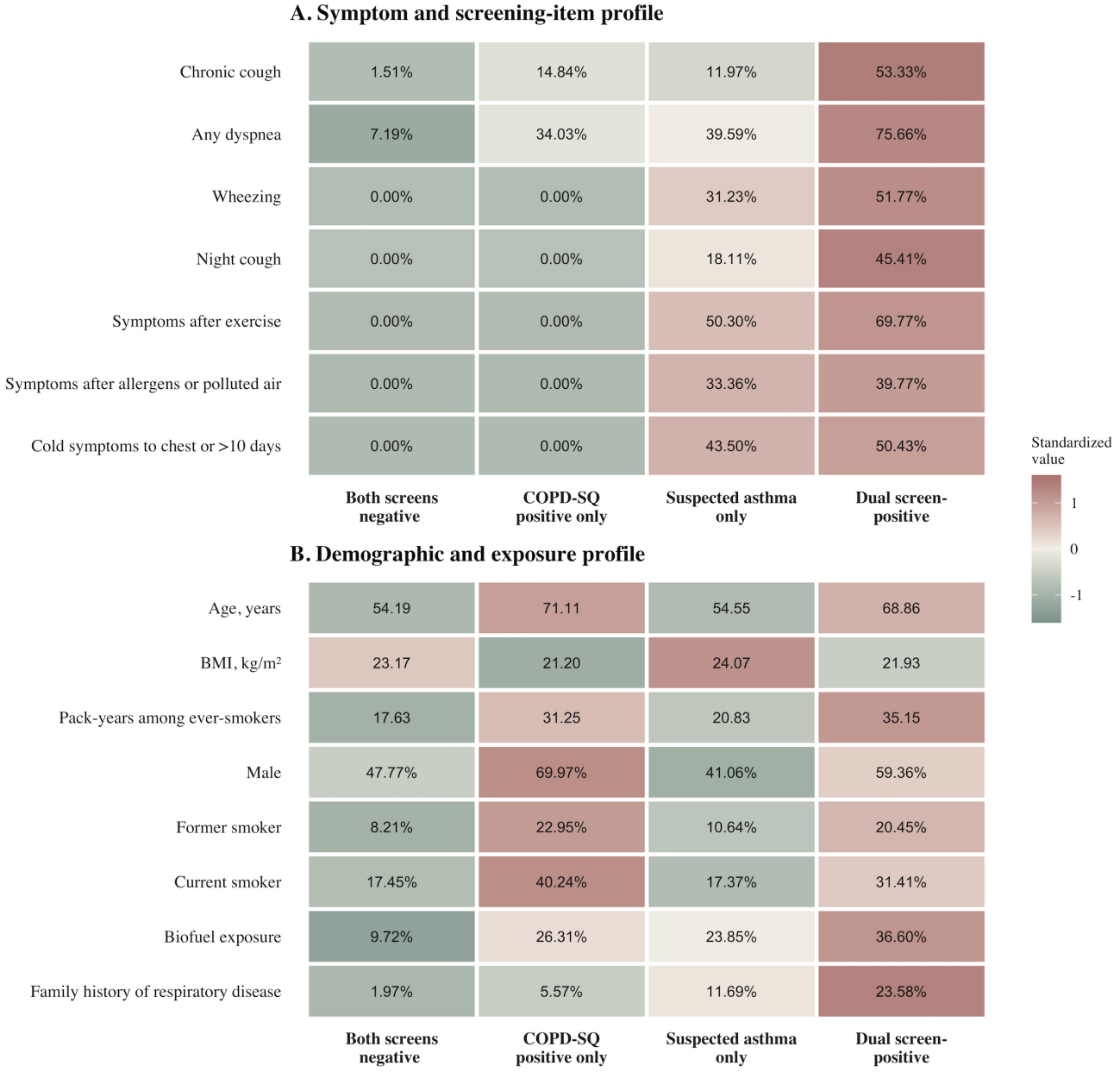
[a]Counts are unweighted observed numbers. Percentages and 95% CIs were estimated using sampling weights adjusted for final analytic inclusion and survey linearized SEs.

The heatmap of standardized group characteristics further illustrated the heterogeneity of the 4 screening-defined risk groups (Figure 3). The COPD-SQ–positive only and dual screen–positive groups showed higher values for age, male sex, and smoking-related indicators than did the both screen-negative and suspected asthma screen–positive only groups. The COPD-SQ–positive only group had the highest mean age and highest proportion of men, whereas the dual screen–positive group had the highest smoking index and the greatest prevalence of biofuel exposure and family history of respiratory disease. The suspected asthma screen–positive only group had the highest BMI and lower smoking exposure than the 2 COPD-SQ–positive groups but still had higher biofuel exposure and a family history of respiratory disease than the both screen–negative group. These patterns suggest that the dual screen–positive group was not simply intermediate between 2 single-positive groups but instead had a combined profile of tobacco exposure, environmental exposure, and family history. Because several characteristics were components of the screening instruments, these patterns should be interpreted descriptively.



XSL•FO
RenderX

**Figure 3.** Weighted profiles of symptoms, screening items, demographic characteristics, and exposures by screening group. Values inside cells are weighted percentages or weighted means. Color intensity represents standardized values across the 4 screening groups, with muted rose indicating relatively higher values and gray sage indicating relatively lower values. Smoking pack-years were calculated among ever-smokers only. Age, BMI, smoking pack-years, chronic cough, dyspnea, biomass smoke exposure, family history of respiratory disease, and all European Community Respiratory Health Survey (ECRHS)–derived symptom items were components of the screening definitions. Accordingly, the between-group patterns for these characteristics are partly definitional and should be interpreted descriptively rather than as independent risk-factor associations. COPD-SQ: Chronic Obstructive Pulmonary Disease Screening Questionnaire.




The weighted proportions with weighted 95% CIs are shown in Table S2 in Multimedia Appendix 1. In a sensitivity analysis using a stricter definition of suspected asthma screen positivity requiring at least 2 positive ECRHS symptom items, the weighted proportion of suspected asthma screen positivity decreased from 22.28% (95% CI 17.26% - 27.30%) to 12.56% (95% CI 9.18%-15.94%), dual screen positivity decreased from 11.31% (95% CI 8.27% - 14.34%) to 7.74% (95% CI 5.47%-10.02%), and positivity on either screening approach decreased from 33.28% (95% CI 26.46% - 40.11%) to 27.13% (95% CI 21.15%-33.12%; see Table S3 in Multimedia Appendix 1). These findings indicate that the estimated overlap was attenuated but not eliminated under the stricter ECRHS definition. In a previous study, a self-reported history of an asthma diagnosis or wheeze symptoms in the preceding 12 months was used to define asthma [4]. Therefore, we redefined suspected asthma screen positivity based on wheeze symptoms and reanalyzed the proportion with suspected asthma screen positivity. The prevalence of wheeze symptom–based suspected asthma was 9.28% (95% CI 6.79% - 11.77%). Further details are provided in Table S4 and Figures S1 and S2 in Multimedia Appendix 1. We also conducted a robustness check to compare weighted screening estimates before and after adjustment for noninclusion in the final analytic sample. The estimates were similar before and after adjustment (see Table S5 in Multimedia Appendix 1).



XSL•FO
RenderX

## Discussion

In this large community-based study of 41,808 adults aged 35 years or older in Xishui County in southwestern China, COPD-SQ and the ECRHS asthma screening questionnaire identified similarly sized but partially overlapping populations. Approximately one-third of participants screened positive on at least one of the 2 approaches, and around 1 in 9 were dual screen-positive. The 4 screening-defined groups also differed in demographic, exposure, and symptom characteristics. These findings indicate that screening questionnaires for COPD and asthma symptoms provide complementary information for identifying unmet health care needs related to these 2 major OLDs in a resource-limited rural setting. In settings where spirometry, trained personnel, and specialist services are limited, a brief digital assessment combining COPD-related risk factors with asthma symptoms may offer a pragmatic first step for stratifying respiratory risk across the community and prioritizing confirmatory evaluation. In the *Discussion* section, we interpreted our results in detail from the following perspectives: the implementation of brief respiratory screening tools in a resource-limited setting, characteristics of different types of screen-positive populations, and implications for the health system.

First, brief respiratory screening tools capture different dimensions of risk and have imperfect diagnostic performance when used alone; therefore, the county-wide digital screening approach generated standardized information on respiratory symptoms, smoking, household biomass smoke exposure, and family history across all townships. Such information has value beyond estimating the proportion of questionnaire-positive residents. It can help characterize the distribution of screening-defined respiratory risk, estimate the potential demand for confirmatory assessment, and identify groups for whom limited diagnostic capacity might be prioritized [23-25]. In mountainous and resource-constrained settings, where residents may face substantial barriers to facility-based assessment, an initial community-based risk-stratification step could reduce the number of people required to travel immediately for spirometry while retaining an opportunity to identify individuals who might otherwise remain outside diagnostic and care pathways. The overlap between the 2 screening instruments also has implications for policymaking. In Chinese primary care, questionnaire-based COPD tools were generally less accurate than simple airflow measurements, while combining questionnaires with airflow assessment improved case detection [26]. Multicountry evidence from Nepal, Peru, and Uganda showed that screening performance varied across settings [27], and the CAPTURE (COPD Assessment in Primary Care to Identify Undiagnosed Respiratory Disease and Exacerbation Risk) study demonstrated modest sensitivity for clinically significant COPD in US primary care [28]. Combining questionnaire information with peak expiratory flow also improved discrimination in resource-limited settings [29]. Earlier work in Mexico similarly proposed a sequential strategy in which questionnaire assessment preceded lung-function testing [30]. Our study addresses a different, upstream question: how COPD-related risk and asthma-related symptoms are distributed and overlap before confirmatory testing. The discordance should therefore be interpreted as complementary risk information. These findings illustrate how screening programs organized exclusively around a single disease label may leave distinct groups of people with respiratory symptoms or exposure-related risks outside the initial assessment pathway [31]. An integrated approach to chronic respiratory risk assessment may therefore be more appropriate than parallel screening programs organized around separate diagnostic labels. Such an approach could link initial risk identification with smoking cessation, reduction of household air pollution, vaccination, confirmatory testing, clinical referral, and subsequent disease-specific management.

Second, the 3 screen-positive subgroups showed different characteristics and should be prioritized in the early detection of COPD and asthma. Participants with dual screen positivity were mostly smokers, were exposed to biofuels, had a family history, or had respiratory symptoms. The COPD-SQ–positivity group was primarily characterized by older age, male predominance, and a higher prevalence of current smoking. The suspected asthma group was characterized by symptoms after exercise, chest symptoms with colds, dyspnea, and wheezing. These characteristics were aligned with previous studies in questionnaire-based screenings. Rural Ugandan evidence highlights both biomass and tobacco consumption as important COPD-related exposures [32]. Smoking and household biomass exposure were associated with clinically defined asthma-COPD overlap [33]. However, several of these characteristics contribute directly to our screening algorithms, so between-group differences are partly determined by classification and should remain descriptive. Clinically characterized asthma-COPD overlap has instead been defined using objective features such as persistent airflow obstruction together with asthma- and COPD-related characteristics [34,35], consistent with broader clinical reviews. Without spirometry or clinical phenotyping, dual positivity is viewed as a potential marker for further assessment rather than a diagnosis or distinct phenotype.

Symptom-based asthma screening via the ECRHS asthma questionnaire is easy to implement but cannot be considered diagnostic. Thus, 2 sensitivity analyses to alter the working definition of suspected asthma were applied, that is, requiring either 2 “yes” answers to the ECRHS items or wheezing symptoms over the past 12 months. Requiring at least 2 ECRHS-derived symptoms reduced the screen-positive proportion, and using wheeze alone produced a still lower estimate of 9.28% (95% CI 6.79% - 11.77%), showing that prevalence depends on the operational threshold, even though a possible working definition of asthma is based on reporting wheezing symptoms over the past 12 months. Symptom-based tools have likewise been used to identify people warranting further asthma assessment without treating questionnaire positivity as a diagnosis; the Childhood Asthma Risk Tool (CHART) provides 1 example in children, although that population is not directly comparable with ours [36]. Ongoing primary-care asthma case-detection research in Uganda similarly evaluates symptom screening together with feasibility, acceptability, and cost-effectiveness [37]. Thus, our 22.28% (95% CI 17.26%-27.30%) estimate should be interpreted as symptom-based screen positivity, not asthma prevalence. The



XSL•FO
RenderX

appropriate threshold will depend on the balance between sensitivity and downstream demand.

The principal implication is that early detection and risk stratification may help organize demand for confirmatory assessment where capacity is constrained. Sequential approaches have previously used questionnaire assessment before physiological testing [26,29,30]. A proactive primary-care e-visit combining COPD screening with smoking-cessation support achieved substantially higher completion of home spirometry than laboratory pulmonary-function testing, while also revealing attrition between screening and definitive testing [38]. In China, an Internet-of-Things–enabled program linked electronic questionnaire screening with spirometry and subsequent COPD management [39]. These examples support evaluating a tiered pathway from risk identification to objective assessment. At the same time, the updated US Preventive Services Task Force (USPSTF) evidence review found no direct evidence that screening asymptomatic adults for COPD improves health outcomes [40]. Our findings therefore inform symptom- and exposure-based prioritization for confirmatory assessment.

An important implication for the health system regarding early detection of CRDs in the primary care setting is that improvements in population health depend more on whether formal diagnoses, treatment, and control follow the initial screening. In the POPMIX series of studies, the COPD-SQ and the ECRHS asthma screening questionnaire were used as the initial screening process. Participants with positive screening results in the intervention arm were invited to the village gathering place for in-depth interviews and physical examinations. In the Undiagnosed COPD and Asthma Population (UCAP) randomized trial, identification of adults with previously undiagnosed symptomatic asthma or COPD followed by pulmonologist-directed treatment reduced respiratory health care use and improved several clinical measures [41]. By contrast, a cluster-randomized CAPTURE implementation trial found that providing screening scores to primary-care clinicians did not significantly improve the prespecified composite of downstream COPD assessment and management actions [42]. These studies suggest that identification alone is insufficient unless connected to an effective care cascade pathway. Digital technologies can support different components, from computed tomography–based deep-learning case finding [43] to electronic questionnaires and remote monitoring. However, scalability should not be equated with effectiveness: asthma e-screening research has identified workflow, clinician, and patient barriers [44], while a pragmatic trial of between-visit digital asthma monitoring showed limited overall improvement in asthma-related quality of life [45]. The initial screening approach in Xishui demonstrates operational feasibility for standardized data collection, but broader scalability, equity, effectiveness, and cost-effectiveness require prospective evaluation.

This study has several strengths. It included a large community population drawn from all townships in Xishui County and collected standardized information on both symptoms and major respiratory exposures. The application of 2 screening approaches allowed us to examine populations that would have been overlooked by a single-disease framework. The QR codes demonstrated the feasibility of standardized respiratory risk assessment across a geographically dispersed population in a mountainous region. The resulting profiles provide locally relevant information for planning confirmatory services and exposure-reduction interventions in a setting that is underrepresented in respiratory epidemiology.

Several limitations should be considered. First, neither spirometry nor clinical assessment was available for the present analysis; thus, the screening categories cannot be interpreted as clinical diagnoses or used to estimate the prevalence of COPD, asthma, or asthma–COPD overlap. Second, several characteristics, including age, BMI, smoking exposure, cough, dyspnea, biomass smoke exposure, family history, and ECRHS-derived symptoms, were components of the screening instruments. Their distributions across the screening-defined groups were therefore partly determined by the classification rules and should not be interpreted as independent risk-factor associations or evidence of disease severity. Third, the primary definition of suspected asthma screen positivity required only one positive ECRHS-derived symptom item and may therefore have limited specificity. Although the main patterns persisted in 2 sensitivity analyses requiring at least 2 positive symptom items and the presence of wheezing symptoms, diagnostic accuracy could not be assessed. Fourth, although the study was designed to provide broad geographic coverage within Xishui County, findings from a county might not be generalizable to other rural or resource-constrained settings. We also did not directly measure travel barriers, spirometry capacity, workforce availability, or the cost of implementing a tiered screening strategy.

In conclusion, community-based screenings may be more informative when they move beyond separate COPD- and asthma-labeled screening pathways and consider complementary patterns of symptoms and exposures. A digitally enabled, integrated risk-stratification strategy may help identify complementary screening-defined risk groups and support tiered allocation of confirmatory spirometry and clinical assessment. In resource-constrained primary care settings, its potential value lies in linking community-based screenings and early detection to spirometry, clinical assessment, referral, preventive interventions, and longitudinal management. Prospective studies should determine whether an integrated, tiered pathway improves diagnostic yield, access, equity, clinical outcomes, and efficiency compared with disease-specific or usual-care approaches.

## Acknowledgments

This work was supported by the People’s Government of Xishui County, Health Bureau of Xishui County, Youth League Committee of Guizhou Medical University, Medieco Group Co. einPaper Team, Finance Bureau of Xishui County, Health



XSL•FO
RenderX

Insurance Bureau of Xishui County, Bureau of Sports and Education of Xishui County, Bureau of Statistics of Xishui County, Xishui Sub-Bureau of the Ecology and Environment Bureau of Zunyi City, Water Authority of Xishui County, Bureau of Housing and Urban-Rural Development of Xishui County, Meteorological Bureau of Xishui County, Center for Disease Control and Prevention of Xishui County (Health Surveillance Station of Xishui County), People's Hospital of Xishui County, Traditional Chinese Medicine Hospital of Xishui County, Maternal and Child Health Care Hospital of Xishui County, and the People's Governments (Offices) of 26 Townships (Subdistricts) and Village Committees of Xishui County.

## Generative AI Use

The authors declare the use of generative AI (GenAI) in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: proofreading and editing, adapting and adjusting emotional tone, reformatting, and quality assessment. The GenAI tools used were ChatGPT 5.6-Thinking (OpenAI) and DeepSeek. Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

## Funding

This study was supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (project number 2024ZD0522300), the National Natural Science Foundation of China (grant number L2424223), AstraZeneca China (project number 2024-HX-22), the CAMS Innovation Fund for Medical Sciences (project number 2023-I2M-2-001), the Non-profit Central Research Institute Fund of the Chinese Academy of Medical Sciences (grant number 2022-ZHCH330-01), and the State Key Laboratory Special Fund (project number 2060204). Additional funding was contributed by the Tencent Sustainable Social Value Inclusive Health Lab (project number SSVPJ202307060001).

## Data Availability

The datasets used or analyzed in this study are available from the principal investigator (CW) upon reasonable request.

## Authors' Contributions

CW (wangchen@pumc.edu.cn) and SC (chensimiao08@icloud.com) are the joint corresponding authors, and Y Liu is the first author. SC, Y Liu, and TB conceived the study idea. Y Liu and SZ contributed to the methodology. Y Liu and SZ contributed to the data analyses. Y Liu, SZ, ZZ, QL, KH, YW, TZ, Y Luo, and XL contributed to data collection and fieldwork. Y Liu wrote the original draft. TB, SC, PG, KH, TY, and CW edited and improved the original draft. CW, SC, and TY provided resources and supervised the study. All authors contributed to the study and agreed to the final submission draft.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Screening classifications, weighted prevalence estimates, and sensitivity analyses.
[DOCX File, 306 KB - i-jmr_v15i1e108291_app1.docx ]

## References

1. Cao Z, He L, Luo Y, et al. Burden of chronic respiratory diseases and their attributable risk factors in 204 countries and territories, 1990-2021: results from the Global Burden of Disease Study 2021. Chin Med J Pulm Crit Care Med 2025 Jun;3(2):100-110. [doi: 10.1016/j.pccm.2025.05.005] [Medline: 40677417]
2. Cherian M, Magner KMA, Whitmore GA, et al. Patient and physician factors associated with symptomatic undiagnosed asthma or COPD. Eur Respir J 2023 Feb;61(2):2201721. [doi: 10.1183/13993003.01721-2022] [Medline: 36328359]
3. Wang C, Xu J, Yang L, et al. Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study. Lancet 2018 Apr 28;391(10131):1706-1717. [doi: 10.1016/S0140-6736(18)30841-9] [Medline: 29650248]
4. Huang K, Yang T, Xu J, et al. Prevalence, risk factors, and management of asthma in China: a national cross-sectional study. Lancet 2019 Aug 3;394(10196):407-418. [doi: 10.1016/S0140-6736(19)31147-X] [Medline: 31230828]
5. Chen S, Kuhn M, Prettner K, et al. The global economic burden of chronic obstructive pulmonary disease for 204 countries and territories in 2020-50: a health-augmented macroeconomic modelling study. Lancet Glob Health 2023 Aug;11(8):e1183-e1193. [doi: 10.1016/S2214-109X(23)00217-6] [Medline: 37474226]
6. Aaron SD, Boulet LP, Reddel HK, Gershon AS. Underdiagnosis and overdiagnosis of asthma. Am J Respir Crit Care Med 2018 Oct 15;198(8):1012-1020. [doi: 10.1164/rccm.201804-0682CI] [Medline: 29756989]
7. Diab N, Gershon AS, Sin DD, et al. Underdiagnosis and overdiagnosis of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2018 Nov 1;198(9):1130-1139. [doi: 10.1164/rccm.201804-0621CI] [Medline: 29979608]



XSL•FO
RenderX

8. Wang C, Qi W, Yang T, et al. The care cascade of chronic obstructive pulmonary disease in China: a cross-sectional study of individual-level data at enrolment into the national 'Happy Breathing' programme. EClinicalMedicine 2024;74:102597. [doi: 10.1016/j.eclinm.2024.102597] [Medline: 39114273]
9. Chen Q, Fan Y, Huang K, et al. Cost-effectiveness of population-based screening for chronic obstructive pulmonary disease in China: a simulation modeling study. Lancet Reg Health West Pac 2024;46:101065. [doi: 10.1016/j.lanwpc.2024.101065] [Medline: 38721063]
10. Zhou YM, Chen SY, Tian J, et al. Development and validation of a Chronic Obstructive Pulmonary Disease Screening Questionnaire in China. Int J Tuberc Lung Dis 2013 Dec;17(12):1645-1651. [doi: 10.5588/ijtld.12.0995] [Medline: 24200283]
11. Beasley R, Semprini A, Mitchell EA. Risk factors for asthma: is prevention possible? Lancet 2015 Sep 12;386(9998):1075-1085. [doi: 10.1016/S0140-6736(15)00156-7] [Medline: 26382999]
12. Liu HM, Huang KW, Wang C. Epidemiology of asthma based on the standard questionnaire [Article in Chinese]. Zhonghua Jie He He Hu Xi Za Zhi 2024 Jan 12;47(1):81-86. [doi: 10.3760/cma.j.cn112147-20231024-00266] [Medline: 38062701]
13. de Marco R, Zanolin ME, Accordini S, et al. A new questionnaire for the repeat of the first stage of the European Community Respiratory Health Survey: a pilot study. Eur Respir J 1999 Nov;14(5):1044-1048. [doi: 10.1183/09031936.99.14510449] [Medline: 10596688]
14. Variations in the prevalence of respiratory symptoms, self-reported asthma attacks, and use of asthma medication in the European Community Respiratory Health Survey (ECRHS). Eur Respir J 1996 Apr;9(4):687-695. [doi: 10.1183/09031936.96.09040687] [Medline: 8726932]
15. Janson C, Anto J, Burney P, et al. The European Community Respiratory Health Survey: what are the main results so far? European Community Respiratory Health Survey II. Eur Respir J 2001 Sep;18(3):598-611. [doi: 10.1183/09031936.01.00205801] [Medline: 11589359]
16. Yuan L, Tao J, Wang J, et al. Global, regional, national burden of asthma from 1990 to 2021, with projections of incidence to 2050: a systematic analysis of the Global Burden of Disease Study 2021. EClinicalMedicine 2025;80:103051. [doi: 10.1016/j.eclinm.2024.103051] [Medline: 39867965]
17. Wang C, Chen S, Shan G, Leng Z, Bärnighausen T, Yang W. Strengthening population medicine to promote public health. Chin Med J (Engl) 2022 May 20;135(10):1135-1137. [doi: 10.1097/CM9.0000000000002221] [Medline: 35788092]
18. Wang C, Liu Y, Huang K, et al. Impact of the Population Medicine Multimorbidity Intervention in Xishui County (POPMIX) on people at high risk for chronic obstructive pulmonary disease: protocol for the POPMIX-COPD cluster randomized controlled trial. JMIR Res Protoc 2026 Feb 18;15:e85597. [doi: 10.2196/85597] [Medline: 41706509]
19. Liu Y, Chen W, Zhang S, et al. Impact of the Population Medicine Multimorbidity Intervention in Xishui County (POPMIX) on people at high risk for chronic obstructive pulmonary disease who experience mental health symptoms: protocol for the POPMIX-MH cluster randomized controlled trial. JMIR Res Protoc 2026 Mar 6;15:e85853. [doi: 10.2196/85853] [Medline: 41791105]
20. Huang K, Tong X, Tang X, et al. Impact of the Population Medicine Multimorbidity Intervention in Xishui County (POPMIX) on suspected asthma patients: protocol for the POPMIX-Asthma cluster-randomized controlled trial. Trials 2026 Apr 11;27(1):301. [doi: 10.1186/s13063-026-09685-5] [Medline: 41964072]
21. Chen S, Huang K, Zheng Z, et al. Impact of the Population Medicine Multimorbidity Intervention In Xishui County (POPMIX) on people at high risk for COPD who smoke: protocol for the POPMIX-Smoking cluster randomised controlled trial. BMJ Open 2026 May 6;16(5):e112179. [doi: 10.1136/bmjopen-2025-112179] [Medline: 42091148]
22. Liu Z, Li YH, Cui ZY, et al. Prevalence of tobacco dependence and associated factors in China: findings from nationwide China Health Literacy Survey during 2018-19. Lancet Reg Health West Pac 2022;24:100464. [doi: 10.1016/j.lanwpc.2022.100464] [Medline: 35538934]
23. Lawes CMM, Thornley S, Young R, et al. Statin use in COPD patients is associated with a reduction in mortality: a national cohort study. Prim Care Respir J 2012 Mar;21(1):35-40. [doi: 10.4104/pcrj.2011.00095] [Medline: 22218819]
24. Guirguis-Blake JM, Senger CA, Webber EM, Mularski RA, Whitlock EP. Screening for chronic obstructive pulmonary disease: evidence report and systematic review for the US Preventive Services Task Force. JAMA 2016 Apr 5;315(13):1378-1393. [doi: 10.1001/jama.2016.2654] [Medline: 27046366]
25. Jordan RE, Adab P, Sitch A, et al. Targeted case finding for chronic obstructive pulmonary disease versus routine practice in primary care (TargetCOPD): a cluster-randomised controlled trial. Lancet Respir Med 2016 Sep;4(9):720-730. [doi: 10.1016/S2213-2600(16)30149-7] [Medline: 27444687]
26. Pan Z, Dickens AP, Chi C, et al. Accuracy and cost-effectiveness of different screening strategies for identifying undiagnosed COPD among primary care patients (≥40 years) in China: a cross-sectional screening test accuracy study: findings from the Breathe Well group. BMJ Open 2021 Sep 23;11(9):e051811. [doi: 10.1136/bmjopen-2021-051811] [Medline: 34556515]
27. Siddharthan T, Pollard SL, Quaderi SA, et al. Discriminative accuracy of chronic obstructive pulmonary disease screening instruments in 3 low- and middle-income country settings. JAMA 2022 Jan 11;327(2):151-160. [doi: 10.1001/jama.2021.23065] [Medline: 35015039]
28. Martinez FJ, Han MK, Lopez C, et al. Discriminative accuracy of the CAPTURE tool for identifying chronic obstructive pulmonary disease in US primary care settings. JAMA 2023 Feb 14;329(6):490-501. [doi: 10.1001/jama.2023.0128] [Medline: 36786790]




XSL•FO
RenderX

29. Checkley W, Yang M, Robertson NM, et al. Population-based screening for chronic obstructive pulmonary disease using the St. George's Respiratory Questionnaire in resource-limited settings. Am J Respir Crit Care Med 2025 May;211(5):779-788. [doi: 10.1164/rccm.202409-1862OC] [Medline: 40068167]
30. Franco-Marina F, Fernandez-Plata R, Torre-Bouscoulet L, et al. Efficient screening for COPD using three steps: a cross-sectional study in Mexico City. NPJ Prim Care Respir Med 2014 May 20;24:14002. [doi: 10.1038/npjpcrm.2014.2] [Medline: 24841708]
31. Leung JM, Sin DD. Asthma-COPD overlap syndrome: pathogenesis, clinical features, and therapeutic targets. BMJ 2017 Sep 25;358:j3772. [doi: 10.1136/bmj.j3772] [Medline: 28947632]
32. van Gemert F, Kirenga B, Chavannes N, et al. Prevalence of chronic obstructive pulmonary disease and associated risk factors in Uganda (FRESH AIR Uganda): a prospective cross-sectional observational study. Lancet Glob Health 2015 Jan;3(1):e44-e51. [doi: 10.1016/S2214-109X(14)70337-7] [Medline: 25539969]
33. Morgan BW, Grigsby MR, Siddharthan T, et al. Epidemiology and risk factors of asthma-chronic obstructive pulmonary disease overlap in low- and middle-income countries. J Allergy Clin Immunol 2019 Apr;143(4):1598-1606. [doi: 10.1016/j.jaci.2018.06.052] [Medline: 30291842]
34. Tommola M, Ilmarinen P, Tuomisto LE, et al. Differences between asthma-COPD overlap syndrome and adult-onset asthma. Eur Respir J 2017 May;49(5):1602383. [doi: 10.1183/13993003.02383-2016] [Medline: 28461298]
35. Uchida A, Sakaue K, Inoue H. Epidemiology of asthma-chronic obstructive pulmonary disease overlap (ACO). Allergol Int 2018 Apr;67(2):165-171. [doi: 10.1016/j.alit.2018.02.002] [Medline: 29551279]
36. Reyna ME, Dai R, Tran MM, et al. Development of a symptom-based tool for screening of children at high risk of preschool asthma. JAMA Netw Open 2022 Oct 3;5(10):e2234714. [doi: 10.1001/jamanetworkopen.2022.34714] [Medline: 36201211]
37. Nantanda R, Ssemata Sentoogo A, Namusoko S, et al. Improving asthma case detection among children and adolescents through clinic-based screening in primary care health facilities in Uganda: a cluster randomised trial. BMJ Open Respir Res 2026 Apr 17;13(1):e003368. [doi: 10.1136/bmjresp-2025-003368] [Medline: 41997611]
38. Dahne J, Player MS, Strange C, et al. Proactive electronic visits for smoking cessation and chronic obstructive pulmonary disease screening in primary care: randomized controlled trial of feasibility, acceptability, and efficacy. J Med Internet Res 2022 Aug 30;24(8):e38663. [doi: 10.2196/38663] [Medline: 36040766]
39. Zhao X, Kang H, An Y, et al. Whole-course management of chronic obstructive pulmonary disease in primary healthcare: an internet of things-enabled prospective cohort study in China. BMJ Open Respir Res 2024 Apr 5;11(1):e001954. [doi: 10.1136/bmjresp-2023-001954] [Medline: 38580439]
40. Webber EM, Lin JS, Thomas RG. Screening for chronic obstructive pulmonary disease: updated evidence report and systematic review for the US Preventive Services Task Force. JAMA 2022 May 10;327(18):1812-1816. [doi: 10.1001/jama.2022.4708] [Medline: 35536261]
41. Aaron SD, Vandemheen KL, Whitmore GA, et al. Early diagnosis and treatment of COPD and asthma - a randomized, controlled trial. N Engl J Med 2024 Jun 13;390(22):2061-2073. [doi: 10.1056/NEJMoa2401389] [Medline: 38767248]
42. Martinez FJ, Yawn BP, Angulo D, et al. Impact of the CAPTURE chronic obstructive pulmonary disease screening tool in U.S. primary care: a cluster-randomized trial. Am J Respir Crit Care Med 2025 May;211(5):789-802. [doi: 10.1164/rccm.202405-0921OC] [Medline: 39715575]
43. Tang LYW, Coxson HO, Lam S, Leipsic J, Tam RC, Sin DD. Towards large-scale case-finding: training and validation of residual networks for detection of chronic obstructive pulmonary disease using low-dose CT. Lancet Digit Health 2020 May;2(5):e259-e267. [doi: 10.1016/S2589-7500(20)30064-9] [Medline: 33328058]
44. MacKinnon M, Moloney M, Bullock E, et al. Implementation of a work-related asthma screening questionnaire in clinical settings: multimethods study. JMIR Form Res 2022 Sep 15;6(9):e37503. [doi: 10.2196/37503] [Medline: 35964327]
45. Rudin RS, Plombon S, Sulca Flores J, et al. Between-visit asthma symptom monitoring with a scalable digital intervention: a randomized clinical trial. JAMA Netw Open 2025 Apr 1;8(4):e256219. [doi: 10.1001/jamanetworkopen.2025.6219] [Medline: 40266619]


## Abbreviations

**CAPTURE:** COPD Assessment in Primary Care to Identify Undiagnosed Respiratory Disease and Exacerbation Risk
**CHART:** Childhood Asthma Risk Tool
**COPD:** chronic obstructive pulmonary disease
**COPD-SQ:** Chronic Obstructive Pulmonary Disease Screening Questionnaire
**CPHS:** China Pulmonary Health Survey
**CRD:** chronic respiratory disease
**ECRHS:** European Community Respiratory Health Survey
**OLD:** obstructive lung disease
**POPMIX:** Population Medicine Multimorbidity Intervention in Xishui County
**UCAP:** Undiagnosed COPD and Asthma Population



XSL•FO
RenderX

**USPSTF:** US Preventive Services Task Force

*Edited by TDA Cardoso; submitted 30.Jul.2026; peer-reviewed by L Lu, S Wang, Z Hao; revised version received 02.Oct.2026; accepted 05.Oct.2026; published 20.Nov.2026.*

*Please cite as:*
*Liu Y, Zhang S, Zheng Z, Huang K, Lai Q, Wang Y, Zhao T, Luo Y, Lyu X, Geldsetzer P, Bärnighausen T, Yang T, Chen S, Wang C COPD-SQ Positivity and Suspected Asthma Screen Positivity From a Representative, Community-Based Sample in Xishui County, Guizhou Province, China: Cross-Sectional Study*
*Interact J Med Res 2026;15:e108291*
*URL: https://www.i-jmr.org/2026/1/e108291*
*doi:10.2196/108291*

© Yuhao Liu, Shiyu Zhang, Zhoutao Zheng, Ke Huang, Qiande Lai, Yingping Wang, Tianying Zhao, Yingchi Luo, Xiangqin Lyu, Pascal Geldsetzer, Till Bärnighausen, Ting Yang, Simiao Chen, Chen Wang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 20.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Value of Blood Count–Derived Inflammatory Markers for Evaluating Psoriasis Severity: Pilot Cross-Sectional Observational Study

Luyuan Wang[1,2*], MD; Xiaorui Zhang[1,2*], MMed; Chuangang Xia[1], MMed; Wenyu Li[3], MD; Qiuju Li[3], MD; Wen Xiao[1], MBBS; Youkun Lin[2,3], MSc

[1]Department of Dermatology and Venereology, Fourth Affiliated Hospital of Guangxi Medical University, Liuzhou, Guangxi, China

[2]First Clinical Medical College, Guangxi Medical University, Nanning, Guangxi, China

[3]Department of Dermatology and Venereology, First Affiliated Hospital of Guangxi Medical University, 6 Shuangyong Road, Nanning, Guangxi, China

[*]these authors contributed equally

**Corresponding Author:**
Youkun Lin, MSc
First Clinical Medical College, Guangxi Medical University, Nanning, Guangxi, China

## Abstract

**Background:** Objective indicators are urgently needed to evaluate and monitor disease progression in patients with psoriasis.

**Objective:** This study aimed to verify the correlations between blood count–derived inflammatory markers and the Psoriasis Area and Severity Index (PASI) among patients with psoriasis and explore the value of applying the PASI in combination with proinflammatory factors.

**Methods:** This was a cross-sectional observational study that enrolled 719 patients from 2 tertiary hospitals. Receiver operating characteristic curve analysis and binary logistic regression models were applied to assess the evaluative power of blood count–derived inflammatory markers and their consistency with the PASI for stratifying psoriasis severity. The association with the PASI and the combination with proinflammatory factors of the blood count–derived inflammatory markers in 60 patients were analyzed. The exploratory association between blood count–derived markers and proinflammatory factors was analyzed using product terms. To ensure robustness, multivariable combined models were evaluated using receiver operating characteristic curves and decision curve analysis. Model performance was further validated via calibration plots and a predictive nomogram, with the decision curve analysis net benefit axis increased to 1.0 for comprehensive visualization.

**Results:** The area under the curve showed that the systemic immune-inflammation index (SII), systemic inflammation response index (SIRI), and aggregate index of systemic inflammation (AISI) were effective in reflecting psoriasis severity and showed advantages in patients with psoriasis complicated by arthritis and cardiovascular metabolic diseases. The comprehensive test showed quite appropriate consistency of the SIRI and PASI in distinguishing severity. The SII, SIRI, and AISI were significantly correlated with interleukin (IL)-6 in lesions (all $P$<.05), and the combinations of these indices with IL-6, IL-1, and IL-17 were also significantly correlated with the PASI (all $P$<.05).

**Conclusions:** Blood count–derived inflammatory markers could better reflect the inflammation of patients with psoriasis. The SII, SIRI, and AISI have important clinical significance in evaluating disease severity. The combination with proinflammatory factors showed an advantage.

***(Interact J Med Res 2026;15:e86454)*** doi:10.2196/86454

**KEYWORDS**

psoriasis; systemic inflammation; interleukins; Psoriasis Area and Severity Index score; PASI score; disease assessment

## Introduction

### Background of Psoriasis and Its Clinical Challenges

Psoriasis is a common chronic inflammatory skin disease that is associated with psoriatic arthritis (PsA), obesity, dyslipidemia, hypertension, type 2 diabetes, and cardiovascular disease [1]. Blood cells interact with proinflammatory factors to regulate the progression of psoriasis. The increase in and prolongation of the systemic inflammatory state in patients may be important determinants of the severity, progression, comorbidities, and outcome of psoriasis. Presently, the treatment model for psoriasis is reactive; therefore, a more proactive and risk-stratified management approach is needed. Early intervention and effective treatment can alleviate the severity of psoriasis by reducing inflammation levels and may also reduce or prevent diseases such as cardiometabolic diseases, as



XSL•FO
RenderX

well as irreversible damage and disability. There is increasing acknowledgment that a single variable is rarely likely to be a good predictor of disease progression [2]; therefore, new parameters are required due to the lack of objective evaluation criteria, insufficient evaluation of chronic inflammation–causing disease, and differences between clinicians.

### Blood Count–Derived Inflammatory Markers: Research Status

Complete blood count parameters have gained recognition as valuable biomarkers for a range of inflammatory conditions, primarily because of their widespread availability and cost-effectiveness. In one study, compared with the healthy control group, patients with psoriasis had higher total white blood cell, neutrophil, and platelet count; neutrophil-to-lymphocyte ratio (NLR); and platelet-to-lymphocyte ratio (PLR) but lower lymphocyte counts. However, there was no significant correlation between the Psoriasis Area and Severity Index (PASI) score and either the NLR or PLR [3]. In another study, the systemic immune-inflammation index (SII) and systemic inflammation response index (SIRI) were powerful prognostic indicators of poor outcomes in patients with cancer and were promising tools for treatment strategy decisions [4,5]. The SIRI and aggregate index of systemic inflammation (AISI) can be useful for the early identification of patients at risk of prolonged hospital stay in open elective thoracic surgery [6]. In another study, the NLR, PLR, monocyte-to-lymphocyte ratio (MLR), SII, SIRI, and AISI were referred to as the blood count–derived inflammatory markers [7]. Both the SII and SIRI have been shown to statistically significantly decrease in patients with psoriasis after only 3 months of treatment with biologics [8,9], and a low correlation between the reduction in PASI scores and PLR, SII, and SIRI values has been observed [9]. The AISI was first reported to be correlated with liver fibrosis severity in psoriasis vulgaris [10]. The AISI, SII, and SIRI have been positively correlated with disease severity in psoriasis [7].

### Research Gaps and Study Rationale

Overall, there are limited studies on the correlation between blood count–derived inflammatory markers and psoriasis. Some studies have had small sample sizes, and there are limited studies that have stratified patients with psoriasis by severity based on PASI score and investigated blood count–derived inflammatory markers. Conventional severity assessment tools, including PASI and body surface area, are insufficient to reflect the underlying pathophysiological features and comprehensive disease course of psoriasis [11]; moreover, the combination of blood count–derived inflammatory markers of psoriasis and proinflammatory factors remains under-studied.

## Methods

### Patients and Data Sources

This study enrolled patients with psoriasis between June 2, 2021, and February 2, 2024, from the Department of Dermatology and Venereology of the First Affiliated Hospital and Fourth Affiliated Hospital of Guangxi Medical University. All patients were diagnosed based on clinical or histopathological manifestations of chronic plaque-type psoriasis. Data on baseline demographics; clinical characteristics; concomitant joint arthritis; and scalp, genital, lower leg, and nail involvement were recorded in the clinic. The disease duration was defined as the period from the first diagnosis of psoriasis by a dermatologist until the present. The inclusion criteria for patients were as follows: (1) patients with chronic plaque-type psoriasis and (2) age of more than 18 years. The PASI was assessed by 3 clinical physicians who had undergone unified training. Exclusion criteria included pregnancy or breastfeeding, acute systemic infection, immunodeficiency, malignant tumors, hematologic disorders, and clinically significant anemia (eg, iron deficiency anemia). Additionally, patients receiving systemic treatments within 1 month preceding enrollment were excluded to ensure a stable baseline and complete blood counts. Standard blood tests were performed for all patients before initiation of treatment, and the results were obtained from the Department of Clinical Laboratory, First Affiliated Hospital and Fourth Affiliated Hospital of Guangxi Medical University. The First Affiliated Hospital cohort served as the derivation set, whereas the Fourth Affiliated Hospital cohort functioned as the independent validation set.

### Ethical Considerations

This study was approved by the ethics committee of the Fourth Affiliated Hospital of Guangxi Medical University (approval KY2024508) and conducted according to the principles of the Declaration of Helsinki. Written informed consent was obtained from all participants.

### Definition of Blood Count–Derived Inflammatory Markers

The NLR was calculated as absolute neutrophil count (× $10^9$/L)/absolute lymphocyte count (× $10^9$/L), the PLR was calculated as absolute platelet count (× $10^9$/L)/absolute lymphocyte count (× $10^9$/L), the MLR was calculated as absolute monocyte count (× $10^9$/L)/absolute lymphocyte count (× $10^9$/L), the SII was calculated as absolute neutrophil count (× $10^9$/L) × absolute platelet count (× $10^9$/L)/absolute lymphocyte count (× $10^9$/L), the SIRI was calculated as absolute neutrophil count (× $10^9$/L) × absolute monocyte count (× $10^9$/L)/absolute lymphocyte count (× $10^9$/L), and the AISI was calculated as absolute neutrophil count (× $10^9$/L) × absolute platelet count (× $10^9$/L) × absolute monocyte count (× $10^9$/L)/absolute lymphocyte count (× $10^9$/L).

### Materials

An antibody against interleukin (IL)-17A (catalog number GB11110-1-50; 1:300) was purchased from ServiceBio, an antibody against IL-1β (cat. no. GB11113-50; 1:300) was purchased from ServiceBio, an antibody against IL-6 (cat. no. YT5348; 1:300) was purchased from Solarbio, and a goat antirabbit secondary antibody (cat. no. 23303-100; 1:200) was purchased from ServiceBio.



XSL•FO
RenderX

### Histomorphometric Analysis and Immunohistochemical Staining

The skin lesions were biopsied at the Department of Dermatology and Venereology of the First Affiliated Hospital of Guangxi Medical University. The lesions were immediately fixed in 4% paraformaldehyde. The samples were embedded in paraffin, and 4 μm thick sections were obtained. Immunohistochemical staining was performed to detect the level of IL-17, IL-1β, and IL-6 from a consecutive sample of 60 patients with psoriasis using the respective antibodies (1:300 dilution) according to the manufacturer's protocol. The histochemistry score (H-score) was adopted for the standardized evaluation of the histological expression of immunohistochemical markers. The specific calculation method was as follows: H-score ($\sum (pi \times i)$ = (percentage of weakly positive cells × 1) + (percentage of moderately positive cells × 2) + (percentage of strongly positive cells × 3), where *i* represents the classification of positive cell grades (strongly positive brownish cells count for 3 points). Positive cells were counted using the ImageJ software. Assessors were blinded to PASI score and blood count–derived inflammatory markers during H-score evaluation.

### Statistical Analysis

Statistical analysis was performed using SPSS (version 25.0; IBM Corp). Statistical descriptions of normal or approximately normal measures were expressed as means and SDs; significantly skewed measures were expressed as medians and IQRs. Furthermore, categorical variables were presented as absolute counts and percentages. Categorical variables were compared using the chi-square test. The optimal cutoff values were determined using the Youden index within the derivation cohort only. Agreement between marker-based and PASI-based classifications in the validation cohort was further assessed using the Cohen κ (with 95% CI) and Pearson chi-square test. The Youden index was calculated as sensitivity − (1 − specificity). The differences between the groups were evaluated using the Mann-Whitney *U* test. Kolmogorov-Smirnov and Shapiro-Wilk tests were used for inspecting the normal distribution. Binary logistic regression analysis was used to further test the results of the Mann-Whitney *U* test. Binary logistic regression included adjustments for BMI, smoking status, and comorbidities (Multimedia Appendix 1). To verify the robustness of these conclusions, a sensitivity analysis was conducted by re-evaluating the inflammatory markers exclusively in a subgroup of patients without any comorbidities (n=571). Due to the skewed distribution and disparate unit scales of the inflammatory markers, natural log-transformation (ln) was applied to improve model stability and interpretability. Extended multivariable models additionally adjusted for study center, disease duration, and PsA. To address potential institutional bias, a stratified receiver operating characteristic (ROC) curve analysis was performed for each hospital cohort, and study center was included as an additional covariate in the sensitivity analysis of the regression model. Comparisons of area under the curve (AUC) values between markers were performed using the DeLong test, with the Benjamini-Hochberg procedure applied to control the false discovery rate (FDR) for multiple comparisons. The Hosmer-Lemeshow test was used to evaluate the goodness of fit of the logistic regression model. Correlations were evaluated using the Spearman correlation coefficient. A *P* value of less than .05 was considered statistically significant. Granular subgroup and correlation analyses were designated as exploratory in nature. A PASI score of 10 or higher vs a PASI score of less than 10 was used as the reference standard for ROC and logistic regression analysis.

## Results

### Baseline Clinical and Demographic Characteristics

The clinical and demographic characteristics of all participants included in the study are shown in Table 1.

Overall, 545 patients were recruited from the Department of Dermatology and Venereology of the First Affiliated Hospital of Guangxi Medical University, and another 174 eligible patients were enrolled from the Department of Dermatology and Venereology of the Fourth Affiliated Hospital of Guangxi Medical University. Among the 719 patients, the lesions involved the scalp in 427 (59.4%) patients, arthritis in 59 (8.2%) patients, the nails in 171 (23.8%) patients, the genital area in 83 (11.5%) patients, and the lower leg in 106 (14.7%) patients. Overall, 148 participants had chronic systemic comorbidities, of whom 97 (65.5%) had cardiometabolic disorders.



XSL•FO
RenderX

**Table .** Characteristics of the patients with psoriasis.

| Characteristic | Values |
|---|---|
| First Affiliated Hospital (n=545) | |
| Gender (woman), n (%) | 141 (25.9) |
| Age (y), mean (SD) | 43.14 (15.76) |
| BMI (kg/m$^2$), median (IQR) | 21.95 (20-23.74) |
| Ever smokers, n (%) | 193 (35.4) |
| Psoriasis duration (y), median (IQR) | 2 (1-6) |
| Specific body part involved, n (%) | |
| Scalp | 304 (55.8) |
| Joint | 48 (8.8) |
| Nails | 121 (22.2) |
| Genitals | 40 (7.3) |
| Lower leg | 79 (14.5) |
| Comorbidity, n (%) | 124 (22.8) |
| PASI[a], median (IQR) | 9 (5.4-16) |
| Fourth Affiliated Hospital (n=174) | |
| Gender (woman), n (%) | 54 (31.0) |
| Age (y), mean (SD) | 49.66 (14.85) |
| BMI (kg/m$^2$), median (IQR) | 24.22 (22.39-27.43) |
| Ever smokers, n (%) | 36 (20.7) |
| Psoriasis duration (y), median (IQR) | 3 (0.66-9) |
| Specific body part involved, n (%) | |
| Scalp | 123 (70.7) |
| Joint | 11 (6.3) |
| Nails | 50 (28.7) |
| Genitals | 43 (24.7) |
| Lower leg | 27 (15.5) |
| Comorbidity, n (%) | 24 (13.8) |
| Treatment, n (%) | |
| Secukinumab | 100 (57.5) |
| Guselkumab | 15 (8.6) |
| Ustekinumab | 9 (5.2) |
| Methotrexate | 31 (17.8) |
| Acitretin | 32 (18.4) |
| Cyclosporin | 22 (12.6) |
| Irregular treatment with secukinumab | 9 (5.2) |
| PASI, median (IQR) | 12.2 (6-21.73) |

[a]PASI: Psoriasis Area and Severity Index.

## Blood Count–Derived Inflammatory Markers and Psoriasis Severity

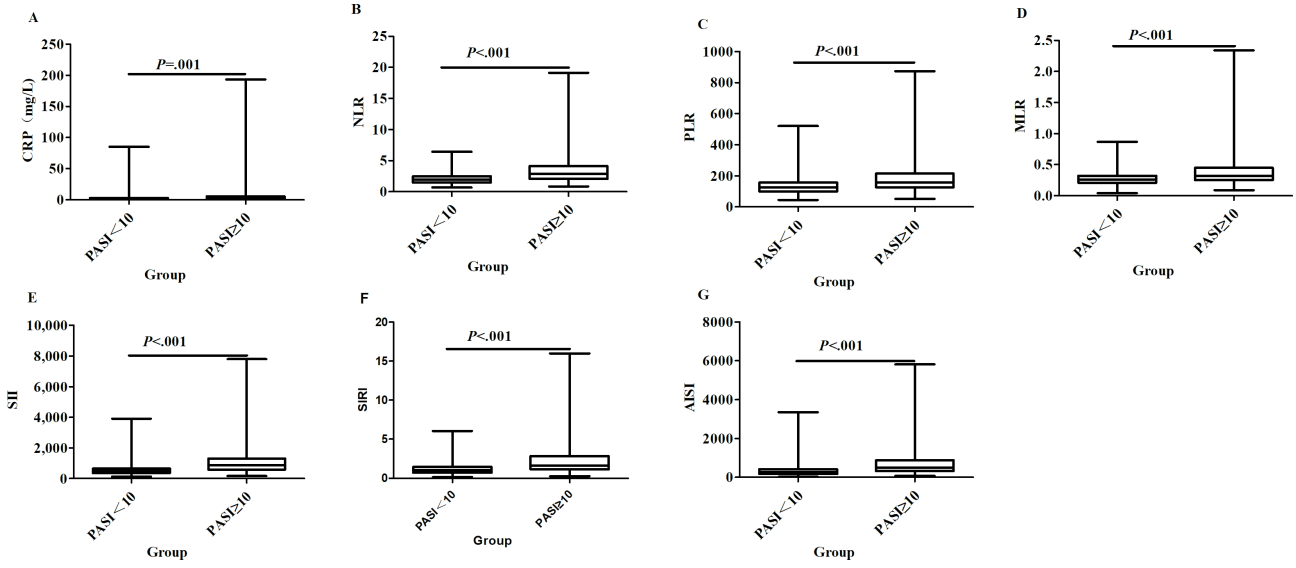
Of the 545 patients with psoriasis in the First Affiliated Hospital, 298 (54.7%) were in the group with a PASI score of less than 10, and 247 (45.3%) were in the group with a PASI score of 10 or higher. NLR, PLR, MLR, SII, SIRI, and AISI were all significantly higher in the group with a PASI score of 10 or higher than in the group with a PASI score of less than 10 ($P$<.001 in all cases; Figure 1). The higher the blood



XSL•FO
RenderX

count–derived inflammatory markers, the higher the psoriasis severity. The results of the ROC curve analyses evaluating the ability of blood count–derived indexes (NLR, PLR, MLR, SII, SIRI, and AISI) to be associated with psoriasis disease severity are presented in Figure 2. The AUC values were 0.58 (95% CI 0.53-0.63) for C-reactive protein (CRP), 0.75 (95% CI 0.70-0.79) for NLR, 0.70 (95% CI 0.66-0.75) for PLR, 0.66 (95% CI 0.62-0.71) for MLR, 0.79 (95% CI 0.75-0.83) for SII, 0.73 (95% CI 0.69-0.78) for SIRI, and 0.77 (95% CI 0.73-0.81) for AISI ($P$<.01 in all cases). The cutoff values were 6.55 for CRP, 2.80 for NLR, 147.85 for PLR, 0.31 for MLR, 662.39 for SII, 1.48 for SIRI, and 443.34 for AISI. The AUC values for NLR, PLR, MLR, SII, SIRI, and AISI were statistically significantly superior to those of CRP (FDR-adjusted $P$<.05 in all cases; Table S1 in Multimedia Appendix 1), suggesting that the blood count–derived indexes were associated with psoriasis severity, although their discriminative performance remained modest. On the basis of the optimal cutoffs from the derivation cohort, the validation cohort showed consistent patterns of association (Table S2 in Multimedia Appendix 1). A consistency test was conducted in 2 groups: those with a PASI score of less than 10 and those with a PASI score of 10 or higher. The comprehensive diagnostic performance, including AUCs with 95% CIs, cutoff values, and Cohen κ analysis, is summarized in Table S3 in Multimedia Appendix 1. The results of the distribution of moderate to severe psoriasis between the SII, SIRI, PLR, and MLR and the PASI score are shown in Table S4 in Multimedia Appendix 1. There were no significant differences in the distribution of moderate to severe psoriasis between the SII, SIRI, PLR, and MLR and the PASI score ($P$>.05 in all cases). Moreover, as shown in Figure S1 in Multimedia Appendix 2, PASI scores were significantly higher among the groups with inflammatory marker values (NLR, MLR, SIRI, SII, and AISI) above the cutoffs compared to those below. However, no statistically significant difference was observed between the CRP and PLR groups. All 719 patients were divided into groups with a PASI score of less than 10 and a PASI score of 10 or higher for binary logistic regression analysis. The Hosmer-Lemeshow test showed a good fit. As shown in Table 2, the NLR, MLR, and SIRI values were higher in the group with a PASI score of 10 or higher. The odds ratios were 1.704 (95% CI 1.483-1.959), 38.132 (95% CI 12.161-119.572), and 1.914 (95% CI 1.605-2.282), respectively ($P$<.001 in all cases). The odds ratios of PLR, SII, AISI, and CRP were close to 1 ($P$<.01 in all cases). In summary, higher NLR, MLR, and SIRI values were independently associated with greater psoriasis severity independent of BMI, smoking status, and comorbidities. To account for scaling effects, log-transformed multivariable analysis was also performed, yielding robust and more interpretable odds ratios (Table S5 in Multimedia Appendix 1). These markers remained independent predictors even after further adjusting for study center, disease duration, and PsA (Table S6 in Multimedia Appendix 1). These associations remained robust after adjusting for study center in the sensitivity analysis (Table S7 in Multimedia Appendix 1). Furthermore, stratified ROC analysis showed consistent predictive performance across both hospital cohorts (Figure S2 in Multimedia Appendix 2).

**Figure 1.** The differential expression of the blood count–derived indexes in patients with moderate and severe psoriasis. AISI: aggregate index of systemic inflammation; CRP: C-reactive protein; MLR: monocyte-to-lymphocyte ratio; NLR: neutrophil-to-lymphocyte ratio; PASI: Psoriasis Area and Severity Index; PLR: platelet-to-lymphocyte ratio; SII: systemic immune-inflammation index; SIRI: systemic inflammation response index.






XSL•FO
RenderX

**Figure 2.** Receiver operating characteristic curve analysis for assessing the performance of the blood count–derived indexes in the prediction of psoriasis activation. AISI: aggregate index of systemic inflammation; AUC: area under the curve; CRP: C-reactive protein; MLR: monocyte-to-lymphocyte ratio; NLR: neutrophil-to-lymphocyte ratio; PLR: platelet-to-lymphocyte ratio; SII: systemic immune-inflammation index; SIRI: systemic inflammation response index.

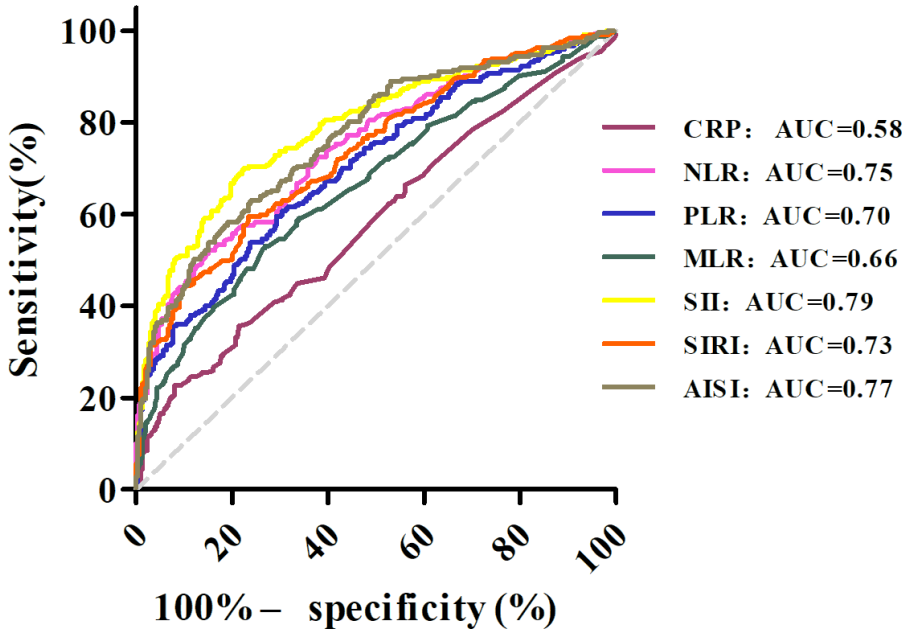



**Table .** Binary logistic regression analysis of associations between blood count–derived inflammatory markers and psoriasis severity.

| | Odds ratio (95% CI) | *P* value | Corrected odds ratio (95% CI) | *P* value |
|---|---|---|---|---|
| NLR[a] | 1.704 (1.483-1.959) | <.001 | 1.683 (1.464 - 1.935) | <.001 |
| PLR[b] | 1.01 (1.008-1.013) | <.001 | 1.011 (1.008 - 1.014) | <.001 |
| MLR[c] | 38.132 (12.161-119.572) | <.001 | 31.815 (10.059-100.630) | <.001 |
| SII[d] | 1.002 (1.001-1.002) | <.001 | 1.002 (1.001-1.002) | <.001 |
| SIRI[e] | 1.914 (1.605-2.282) | <.001 | 1.878 (1.574-2.242) | <.001 |
| AISI[f] | 1.002 (1.002-1.003) | <.001 | 1.002 (1.002-1.003) | <.001 |
| CRP[g] | 1.024 (1.009-1.038) | .001 | 1.021 (1.007-1.036) | .004 |

[a]NLR: neutrophil-to-lymphocyte ratio.

[b]PLR: platelet-to-lymphocyte ratio.

[c]MLR: monocyte-to-lymphocyte ratio.

[d]SII: systemic immune-inflammation index.

[e]SIRI: systemic inflammation response index.

[f]AISI: aggregate index of systemic inflammation.

[g]CRP: C-reactive protein.

## Subgroup Analysis of Blood Count–Derived Inflammatory Markers

The NLR, PLR, MLR, SII, SIRI, and AISI values in the group with joint involvement were significantly higher than those in the group without joint involvement, and the MLR, SIRI, and AISI values were higher in the groups with scalp and nail involvement. The NLR, PLR, MLR, SII, SIRI, and AISI values did not show significant increases in the groups with calf and genital involvement (Table S8 in Multimedia Appendix 1). The NLR, MLR, SII, SIRI, and AISI values of patients with combined cardiovascular or metabolic diseases were significantly higher and statistically significant (Table S9 in Multimedia Appendix 1).

## Combined Analysis of Markers and Cytokines

Immunohistochemical staining of IL-1, IL-6, and IL-17 was performed on patients with psoriasis at the First Affiliated Hospital. Demographic and clinical comparison confirmed that this subcohort was representative of the full cohort (Table S10 in Multimedia Appendix 1). There were positive expressions around keratinocytes and superficial dermal inflammatory cells, as shown in Figure S3 in Multimedia Appendix 2. IL-1 and IL-17 were not expressed in the normal skin around the mole.



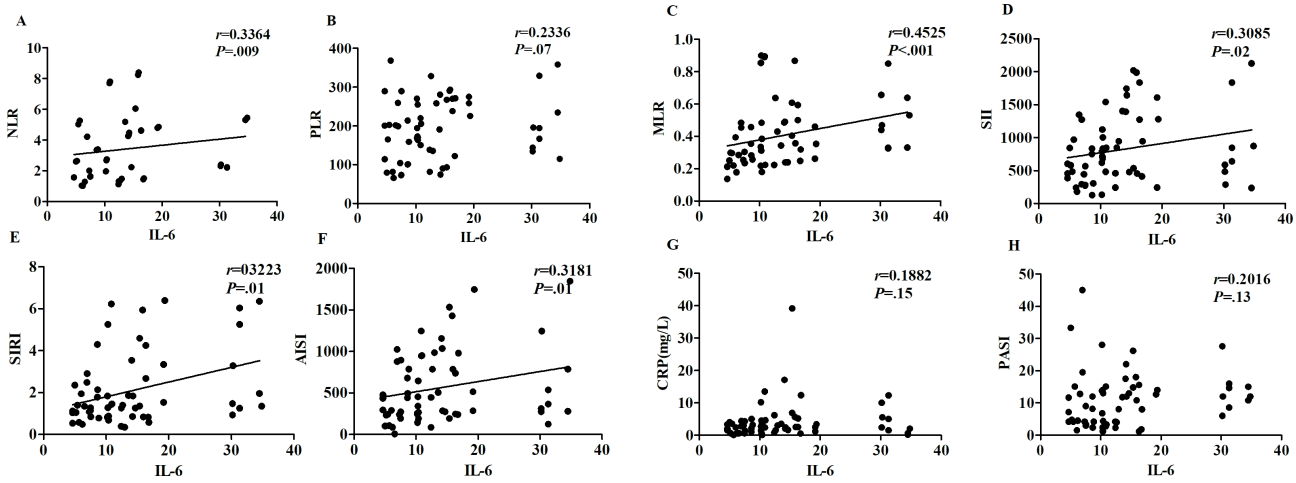
XSL•FO
RenderX

However, by calculating the H-score, it was found that IL-17 and IL-1 had no significant correlation with the PASI score, as shown in Figures S4 and S5 in Multimedia Appendix 2.

Moreover, there was no statistically significant correlation between the blood count–derived index and IL-1 and IL-17. The correlations between NLR, MLR, SII, SIRI, and AISI and IL-6 in the skin lesions were statistically significant, with correlation coefficients of 0.3364 (*P*=.009), 0.4525 (*P*<.001), 0.3085 (*P*=.02), 0.3223 (*P*=.01), and 0.3181 (*P*=.01), respectively (Figure 3). There was no correlation between IL-6 and PASI score in the skin lesions (*P*=.13). There was no statistically significant correlation between PLR and CRP (*P*=.07 and .15, respectively).

We conducted an exploratory analysis combining IL-6, IL-1, and IL-17 with the blood count–derived indexes and found that the blood count–derived indexes combined with IL-6 in the skin lesions of patients with psoriasis had a significant correlation with the PASI score, as shown in Table 3. The correlation coefficients of the combination of IL-6 with NLR, PLR, MLR, SII, SIRI, and AISI with PASI scores were 0.66, 0.46, 0.40, 0.64, 0.56, and 0.56, respectively. The combination of IL-1 with NLR and SII and the combination of IL-17 with NLR and SII were significantly correlated with PASI scores. The correlation coefficients were 0.48, 0.50, 0.49, and 0.48, respectively. The combined multivariable model (eg, SIRI+IL-6) showed stronger associations with disease severity (AUC=0.845) and a more favorable clinical net benefit in the decision curve analysis (Figure S6 in Multimedia Appendix 2). To achieve comprehensive visual evaluation of clinical net benefit, the DCA net benefit axis was raised to 1.0 in the present analysis. Calibration analysis and a functional clinical nomogram with an individual case example further confirmed high reliability and facilitated bedside risk assessment. As summarized in the integrated correlation matrix (Figure S7 in Multimedia Appendix 2), the associations remained highly significant after FDR adjustment.

**Figure 3.** Correlation between blood count–derived indexes and interleukin (IL)-6. AISI: aggregate index of systemic inflammation; CRP: C-reactive protein; MLR: monocyte-to-lymphocyte ratio; NLR: neutrophil-to-lymphocyte ratio; PASI: Psoriasis Area and Severity Index; PLR: platelet-to-lymphocyte ratio; SII: systemic immune-inflammation index; SIRI: systemic inflammation response index.






XSL•FO
RenderX

**Table .** The correlation between the combination of blood count–derived inflammatory markers with interleukin (IL)-6, IL-1, and IL-17 and the Psoriasis Area and Severity Index (PASI).

| Interleukin and index it was combined with | Correlation with the PASI | |
|---|---|---|
| | *r* | *P* value |
| IL-6 | | |
| NLR[a] | 0.66 | <.001 |
| PLR[b] | 0.46 | .02 |
| MLR[c] | 0.40 | .05 |
| SII[d] | 0.64 | .001 |
| SIRI[e] | 0.56 | .003 |
| AISI[f] | 0.56 | .004 |
| IL-1 | | |
| NLR | 0.48 | <.001 |
| PLR | 0.38 | .003 |
| MLR | 0.22 | .09 |
| SII | 0.50 | <.001 |
| SIRI | 0.31 | .01 |
| AISI | 0.34 | .007 |
| IL-17 | | |
| NLR | 0.49 | <.001 |
| PLR | 0.44 | .001 |
| MLR | 0.23 | .08 |
| SII | 0.48 | <.001 |
| SIRI | 0.31 | .02 |
| AISI | 0.34 | .008 |

[a]NLR: neutrophil-to-lymphocyte ratio.

[b]PLR: platelet-to-lymphocyte ratio.

[c]MLR: monocyte-to-lymphocyte ratio.

[d]SII: systemic immune-inflammation index.

[e]SIRI: systemic inflammation response index.

[f]AISI: aggregate index of systemic inflammation.

## Discussion

Psoriasis is driven by combined autoimmune and autoinflammatory responses, in which IL-1, IL-17, and IL-6 form an inflammatory cascade by regulating T helper 17 cells, neutrophils, and regulatory T cells [12,13]. Traditional clinical assessments such as PASI and body surface area inadequately reflect systemic inflammation, whereas cytokine-based biomarkers are costly and fail to clearly distinguish current severity from future risk [2,11]. Given the high cost and time consumption associated with specialized cytokine biomarkers, PLR, NLR, MLR, SII, SIRI, and AISI, which could be easily derived as part of a routine complete blood count, have been used in chronic inflammatory diseases, including psoriasis and psoriasis comorbidity [7,14,15]. These indicators of inflammation include neutrophil, platelet, monocyte, or lymphocyte count. Recent reports have indicated that oxidative stress, granular components, and neutrophil extracellular traps are linked to psoriasis initiation and maintenance [16]. The abnormality of the immune system in psoriasis may overactivate platelets, contributing to the disease and its cardiovascular comorbidities, with antiplatelet therapies potentially being effective [17]. A genome-wide study found that a high monocyte and neutrophil count and elevated NLR and PLR increase psoriasis risk [18], whereas patients with psoriasis vulgaris had a lower lymphocyte count than controls [3]. The excessive proinflammatory immune response in psoriasis drives the proliferation of neutrophils, platelets, and monocytes, accompanied by relative lymphocyte depletion.

Previous studies have reported elevated NLR and PLR in patients with psoriasis, although correlations with the PASI score are inconsistent [3,19]. A cohort study linked NLR to baseline PASI score and noncalcified coronary plaque burden [20]. Furthermore, SII was elevated in patients with higher PASI



XSL•FO
RenderX

scores and nail or genital involvement [21]. Population-based National Health and Nutrition Examination Survey data further demonstrated positive associations between NLR, PLR, SII, and SIRI and psoriasis risk [22,23], and higher MLR and SIRI were positively associated with all-cause mortality in adults with psoriasis [23]. PLR and CRP were associated with the diagnosis of PsA, whereas SII and SIRI were not. Patients with higher PLR and SII scores were more resistant to treatment with conventional systemic agents, although no significant differences in these markers were found among psoriasis, PsA, and controls [15]. A recent meta-analysis of 36 studies confirmed significantly higher NLR, PLR, SII, and MLR in patients with psoriasis vs controls, with positive correlations between both NLR and PLR and the PASI score [24]. Inconsistent findings across studies may stem from different cohorts or populations, requiring further large-scale research. In this study, the NLR, PLR, SII, SIRI, and AISI of patients with psoriasis were higher in the group with a PASI score of 10 or higher (Figure 1), with all AUC values exceeding 0.7, supporting their potential role as exploratory severity markers (Figure 2). It was found that there was no statistically significant difference in distribution between the SII, SIRI, PLR, and MLR results and the PASI score in diagnosing moderate to severe psoriasis (Table S4 in Multimedia Appendix 1). Among them, the *P* values of the SII and AISI were .05 in both cases, which might be related to the relatively small number of patients. The higher cutoff values for the NLR, MLR, and SIRI, the higher PASI scores (Figure S1 in Multimedia Appendix 2) and the higher severity level of patients with psoriasis. The independent predictive value of these markers remained robust across different modeling approaches (Table 2). Importantly, sensitivity analysis restricted to patients without comorbidities (n=571) yielded highly consistent odds ratios, confirming that these findings were independent of systemic confounding bias (Table S11 in Multimedia Appendix 1). Comorbidity was modeled as a binary covariate; its detailed classification and modeling framework are provided in Tables S11 and S12 in Multimedia Appendix 1. On the basis of the above analysis, these findings suggest that the SIRI is associated with PASI scores and may help stratify moderate to severe psoriasis, although discriminative accuracy is not sufficiently robust for stand-alone clinical use. The robustness of these markers was further confirmed using log-transformed models, which effectively mitigated the influence of outliers and unit scaling issues. In one study, the SII was higher in patients with PsA than in those without arthritis [25]. Another study mentioned that the SII and SIRI had no significant correlation with the diagnosis of PsA, whereas SII and SIRI values were higher in patients with psoriasis involving the nails [8]. However, subgroup analyses in our study further revealed significantly higher blood count–derived indexes in patients with joint, scalp, or nail involvement, including those with cardiometabolic comorbidities (Tables S3 and S4 in Multimedia Appendix 1). These observations support the notion that blood count–derived inflammatory markers reflect systemic inflammatory burden beyond cutaneous involvement alone. The small sample size of patients with special site involvement and cardiometabolic comorbidities may introduce selection bias; future multicenter studies with larger populations are warranted to verify these findings. The PASI score is the most widely accepted and routinely used clinical index for assessing psoriasis severity. In this study, the PASI score was used as a benchmark outcome for ROC and logistic regression analyses, not as a perfect or gold-standard reflection of overall disease activity.

Currently, standardized objective biomarkers for psoriasis evaluation remain limited, and few studies have explored the combined application of blood count–derived inflammatory markers and proinflammatory cytokines. It was observed that there were positive expressions of IL-6, IL-1, and IL-17 to varying degrees around keratinocytes and superficial dermal inflammatory cells (Figure S2 in Multimedia Appendix 2). Notably, although IL-17 and IL-1 showed no significant correlation with the PASI score in the dataset, the NLR, MLR, SII, SIRI, and AISI were significantly associated with lesional IL-6 expression. This suggests that blood count–derived inflammatory markers may indirectly reflect IL-6–driven myeloid and lymphoid imbalance in psoriasis lesions. In this study, by combining IL-6, IL-1, and IL-17 with the blood count–derived inflammatory markers, it was found that there was a significant correlation with the PASI score, especially in combination with IL-6 (Table 3).

Limitations included the reliance on 2 regional centers, potential residual or unmeasured confounders, limited generalizability due to enrolling only patients requiring systemic therapy, bias from using normal perilesional skin as a control due to local microenvironmental factors, and lack of longitudinal data and external validation. Furthermore, while primary findings were validated using FDR correction, the multidimensional subgroup analyses remain exploratory and require further validation. The cytokine analysis was limited to a small subset as they may be prone to overfitting and lack sufficient generalizability; these exploratory findings should be interpreted cautiously. All these limitations should be addressed in the future.

In conclusion, blood count–derived inflammatory markers were significantly associated with the PASI score in psoriasis and correlated with lesional IL-6 expression. Moreover, the combination of blood count–derived inflammatory markers and proinflammatory factors may serve as a choice for accurately and comprehensively evaluating psoriasis severity and monitoring treatment response in the future, providing new ideas for multidimensional evaluation. As low-cost, accessible alternatives to expensive cytokine profiling, these markers support a shift from reactive to proactive, risk-stratified psoriasis management, aiding early identification of high–inflammatory burden patients. The link between blood inflammatory indexes and lesional IL-6 signaling confirms their biological relevance as surrogates of tissue-level immune dysregulation. The combined application of these markers complements traditional scores, advancing precision dermatology for more objective, individualized disease evaluation and improved long-term outcomes.



XSL•FO
RenderX

## Acknowledgments

The authors would like to express their gratitude to all participants in this study.

## Funding

This work was supported by grants from the National Natural Science Foundation of China (82260620), Science and Technology Program of Liuzhou (2022CAC0227), and Key Research and Development Program of Guangxi (2022AB08042).

## Data Availability

Data are available from the corresponding author on reasonable request.

## Authors' Contributions

WX is the co–corresponding author and can be reached via email at 196821xw@sina.com or phone at 86 0772-3815455. LW conceived and designed the study. XZ and CX performed the data collection and statistical analysis. WL and QL contributed to the methodology and formal analysis. LW drafted the original manuscript. WX and YL revised the manuscript critically for important intellectual content. WX and YL were responsible for supervision, resources, and project administration. All authors have read and approved the final manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Inflammatory markers, diagnostic performance, and related analyses in patients with psoriasis.
[DOCX File, 43 KB - i-jmr_v15i1e86454_app1.docx ]

Multimedia Appendix 2
Analyses of psoriasis inflammation, blood count-derived marker performance, cytokine expression and correlations, and multivariable predictive model evaluation.
[DOCX File, 2338 KB - i-jmr_v15i1e86454_app2.docx ]

## References

1. Griffiths CEM, Armstrong AW, Gudjonsson JE, Barker J. Psoriasis. Lancet 2021 Apr 3;397(10281):1301-1315. [doi: 10.1016/S0140-6736(20)32549-6] [Medline: 33812489]
2. Ramessur R, Corbett M, Marshall D, et al. Biomarkers of disease progression in people with psoriasis: a scoping review. Br J Dermatol 2022 Oct;187(4):481-493. [doi: 10.1111/bjd.21627] [Medline: 35482474]
3. Wang WM, Wu C, Gao YM, Li F, Yu XL, Jin HZ. Neutrophil to lymphocyte ratio, platelet to lymphocyte ratio, and other hematological parameters in psoriasis patients. BMC Immunol 2021 Sep 26;22(1):64. [doi: 10.1186/s12865-021-00454-4] [Medline: 34565327]
4. Qi Q, Zhuang L, Shen Y, et al. A novel systemic inflammation response index (SIRI) for predicting the survival of patients with pancreatic cancer after chemotherapy. Cancer 2016 Jul 15;122(14):2158-2167. [doi: 10.1002/cncr.30057] [Medline: 27152949]
5. Hu B, Yang XR, Xu Y, et al. Systemic immune-inflammation index predicts prognosis of patients after curative resection for hepatocellular carcinoma. Clin Cancer Res 2014 Dec 1;20(23):6212-6222. [doi: 10.1158/1078-0432.CCR-14-0442] [Medline: 25271081]
6. Paliogiannis P, Ginesu GC, Tanda C, et al. Inflammatory cell indexes as preoperative predictors of hospital stay in open elective thoracic surgery. ANZ J Surg 2018 Jun;88(6):616-620. [doi: 10.1111/ans.14557] [Medline: 29687547]
7. Tiucă OM, Morariu SH, Mariean CR, Tiucă RA, Nicolescu AC, Cotoi OS. Impact of blood-count-derived inflammatory markers in psoriatic disease progression. Life (Basel) 2024 Jan 12;14(1):114. [doi: 10.3390/life14010114] [Medline: 38255729]
8. Tamer F, Edek YC, Aksakal AB. Effect of treatment with biologic agents on the novel inflammatory biomarkers systemic immune inflammation index and systemic inflammation response index for psoriasis. Dermatol Pract Concept 2024 Jan 1;14(1):e2024065. [doi: 10.5826/dpc.1401a65] [Medline: 38364414]
9. Kulakli S, Oguz ID, Aksan B. Could blood cell-based inflammatory markers be used to monitor response to biologic therapy in psoriasis? Sisli Etfal Hastan Tip Bul 2023;57(4):536-542. [doi: 10.14744/SEMB.2023.43569] [Medline: 38268655]
10. Tiucă OM, Morariu SH, Mariean CR, Tiucă RA, Nicolescu AC, Cotoi OS. Predictive performances of blood-count-derived inflammatory markers for liver fibrosis severity in psoriasis vulgaris. Int J Mol Sci 2023 Nov 29;24(23):16898. [doi: 10.3390/ijms242316898] [Medline: 38069218]



XSL•FO
RenderX

11. Christophers E, van de Kerkhof PCM. Severity, heterogeneity and systemic inflammation in psoriasis. J Eur Acad Dermatol Venereol 2019 Apr;33(4):643-647. [doi: 10.1111/jdv.15339] [Medline: 30924245]
12. Cai Y, Xue F, Quan C, et al. A critical role of the IL-1β-IL-1R signaling pathway in skin inflammation and psoriasis pathogenesis. J Invest Dermatol 2019 Jan;139(1):146-156. [doi: 10.1016/j.jid.2018.07.025] [Medline: 30120937]
13. Xu H, Liu J, Niu M, et al. Soluble IL-6R-mediated IL-6 trans-signaling activation contributes to the pathological development of psoriasis. J Mol Med (Berl) 2021 Jul;99(7):1009-1020. [doi: 10.1007/s00109-021-02073-3] [Medline: 33835216]
14. Islam MM, Satici MO, Eroglu SE. Unraveling the clinical significance and prognostic value of the neutrophil-to-lymphocyte ratio, platelet-to-lymphocyte ratio, systemic immune-inflammation index, systemic inflammation response index, and delta neutrophil index: an extensive literature review. Turk J Emerg Med 2024;24(1):8-19. [doi: 10.4103/tjem.tjem_198_23] [Medline: 38343523]
15. Sugimoto E, Matsuda H, Shibata S, et al. Impact of pretreatment systemic inflammatory markers on treatment persistence with biologics and conventional systemic therapy: a retrospective study of patients with psoriasis vulgaris and psoriatic arthritis. J Clin Med 2023 Apr 21;12(8):3046. [doi: 10.3390/jcm12083046] [Medline: 37109382]
16. Chiang CC, Cheng WJ, Korinek M, Lin CY, Hwang TL. Neutrophils in psoriasis. Front Immunol 2019;10:2376. [doi: 10.3389/fimmu.2019.02376] [Medline: 31649677]
17. Jiang Z, Jiang X, Chen A, He W. Platelet activation: a promoter for psoriasis and its comorbidity, cardiovascular disease. Front Immunol 2023;14:1238647. [doi: 10.3389/fimmu.2023.1238647] [Medline: 37654493]
18. Zhou G, Ren X, Tang Z, et al. Exploring the association and causal effect between white blood cells and psoriasis using large-scale population data. Front Immunol 2023;14:1043380. [doi: 10.3389/fimmu.2023.1043380] [Medline: 36865550]
19. Ye JH, Zhang Y, Naidoo K, Ye S. Neutrophil-to-lymphocyte ratio and platelet-to-lymphocyte ratio in psoriasis: a systematic review and meta-analysis. Arch Dermatol Res 2024 Feb 8;316(3):85. [doi: 10.1007/s00403-024-02823-6] [Medline: 38329632]
20. Dey AK, Teague HL, Adamstein NH, et al. Association of neutrophil-to-lymphocyte ratio with non-calcified coronary artery burden in psoriasis: findings from an observational cohort study. J Cardiovasc Comput Tomogr 2021;15(4):372-379. [doi: 10.1016/j.jcct.2020.12.006] [Medline: 33390348]
21. Dincer Rota D, Tanacan E. The utility of systemic-immune inflammation index for predicting the disease activation in patients with psoriasis. Int J Clin Pract 2021 Jun;75(6):e14101. [doi: 10.1111/ijcp.14101] [Medline: 33619821]
22. Ma R, Cui L, Cai J, et al. Association between systemic immune inflammation index, systemic inflammation response index and adult psoriasis: evidence from NHANES. Front Immunol 2024;15:1323174. [doi: 10.3389/fimmu.2024.1323174] [Medline: 38415255]
23. Zhao Y, Yang XT, Bai YP, Li LF. Association of complete blood cell count-derived inflammatory biomarkers with psoriasis and mortality. Clin Cosmet Investig Dermatol 2023;16:3267-3278. [doi: 10.2147/CCID.S437936] [Medline: 38021430]
24. Liu YC, Chuang SH, Chen YP, Shih YH. Associations of novel complete blood count-derived inflammatory markers with psoriasis: a systematic review and meta-analysis. Arch Dermatol Res 2024 May 24;316(6):228. [doi: 10.1007/s00403-024-02994-2] [Medline: 38787437]
25. Yorulmaz A, Hayran Y, Akpinar U, Yalcin B. Systemic immune-inflammation index (SII) predicts increased severity in psoriasis and psoriatic arthritis. Curr Health Sci J 2020;46(4):352-357. [doi: 10.12865/CHSJ.46.04.05] [Medline: 33717509]


## Abbreviations

**AISI:** aggregate index of systemic inflammation
**AUC:** area under the curve
**CRP:** C-reactive protein
**FDR:** false discovery rate
**H-score:** histochemistry score
**IL:** interleukin
**MLR:** monocyte-to-lymphocyte ratio
**NLR:** neutrophil-to-lymphocyte ratio
**PASI:** Psoriasis Area and Severity Index
**PLR:** platelet-to-lymphocyte ratio
**PsA:** psoriatic arthritis
**ROC:** receiver operating characteristic
**SII:** systemic immune-inflammation index
**SIRI:** systemic inflammation response index



XSL•FO
RenderX

*Edited by M Balcarras; submitted 24.Oct.2025; peer-reviewed by A Dervisevic, C Cifuentes-Gonzalez, ND Ogut, R Aci; revised version received 20.Apr.2026; accepted 20.Apr.2026; published 14.May.2026.*

*Please cite as:*
*Wang L, Zhang X, Xia C, Li W, Li Q, Xiao W, Lin Y*
*Value of Blood Count–Derived Inflammatory Markers for Evaluating Psoriasis Severity: Pilot Cross-Sectional Observational Study*
*Interact J Med Res 2026;15:e86454*
*URL: https://www.i-jmr.org/2026/1/e86454*
*doi:10.2196/86454*

© Luyuan Wang, Xiaorui Zhang, Chuangang Xia, Wenyu Li, Qiuju Li, Wen Xiao, Youkun Lin. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 14.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Mapping the Global Surge in Postoperative Sleep Research From 2014 to 2024: Bibliometric Analysis

Xin Wei[1], BD; Lijuan Fu[1], MD; Wencai Jiang[1], MD; Xianjie Zhang[1], MD; Rui Zhou[2], MD

[1]Department of Anesthesiology, Deyang People's Hospital, Deyang, China

[2]Department of Anesthesiology and Perioperative Medicine, Shanghai Fourth People's Hospital, No 1279, Sanmen Road, Hongkou District, Shanghai, China

**Corresponding Author:**
Rui Zhou, MD
Department of Anesthesiology and Perioperative Medicine, Shanghai Fourth People's Hospital, No 1279, Sanmen Road, Hongkou District, Shanghai, China

## Abstract

**Background:** Postoperative sleep is closely associated with recovery among patients undergoing surgery.

**Objective:** This study aimed to analyze the research status and developmental trends in postoperative sleep between 2014 and 2024.

**Methods:** Publications were retrieved from the Web of Science Core Collection. Microsoft Excel and VOSviewer were used to analyze the papers in terms of publication trends, countries, institutions, authors, journals, and keywords.

**Results:** A total of 964 papers were obtained for the bibliometric analysis. The number of publications on this topic has increased gradually over the last 10 years. Zhu Junchao was the most prolific author in the field, and Chung Frances had the most citations. China had the most publications, followed by the United States. Scientific institutions in China, such as China Medical University and Capital Medical University, have led the way in terms of publication numbers. A total of 40 journals have published at least 5 papers. *BMC Anesthesiology*, with 19 publications, ranked first in publication count. The papers published in the *British Journal of Anaesthesia*, *Journal of Clinical Anesthesia*, *Anesthesia and Analgesia*, *Journal of Pain*, and *Journal of Sleep Medicine* had higher citation counts on average. The high-frequency keywords were “sleep quality,” “postoperative pain,” “quality of life,” and “surgery,” while “lung cancer,” “enhanced recovery after surgery,” “breast cancer,” and “dexamethasone” emerged as new topics in this area.

**Conclusions:** There has emerged a large body of literature on postoperative sleep over the past 10 years. Authors and organizations from China are leading contributors, followed by those from the United States. Anesthesiology is a critical discipline in this field. Postoperative pain is closely related to postoperative sleep and has become a major research focus. Recent studies have mainly focused on lung cancer and breast cancer surgeries. Enhanced recovery after surgery has become an emerging keyword.

***(Interact J Med Res 2026;15:e86086)*** doi:10.2196/86086

**KEYWORDS**

bibliometric analysis; postoperative sleep; recovery quality; postoperative pain; enhanced recovery after surgery

## Introduction

Postoperative sleep shows a close relationship with the quality of recovery, hormone secretion, metabolism, cognitive function, and autonomic nervous activity [1]. Maintenance of normal sleep after surgery is closely related to postoperative rehabilitation [2]. However, patients undergoing surgery may experience a degree of impact on their sleep quality [3]. Anesthesia has the potential to disrupt the sleep-wake timing system [4,5], as most general anesthetics are either N-methyl-D-aspartate receptor antagonists or γ-aminobutyric acid agonists, and activation of these receptors affects clock gene expression [6]. Glucocorticoids, synthetic analogs of endogenous cortisol, are sometimes administered during surgery and have strong effects on the molecular clock [7]. In addition, the grade and duration of surgery have been demonstrated to determine the degree of postoperative sleep disorders [8,9]. Thus, normal sleep after surgery may be difficult to maintain because of the multifaceted effects of surgery and anesthesia on sleep during the early postoperative period.

Postoperative sleep disorders have received increasing attention from researchers [10]. More than 1900 publications were recorded in the Web of Science (WoS) Core Collection from 2014 to 2023. Despite the increasing number of studies, a comprehensive and systematic overview of the global research status, developmental trends, and research hot spots in postoperative sleep remains lacking. Previous studies have mainly focused on specific clinical issues such as the correlation



XSL•FO
RenderX

between sleep quality and postoperative pain or the effect of a single intervention on sleep improvement. However, there is a lack of macrolevel analyses summarizing the research landscape, identifying collaborative networks, and predicting future research directions in this field.

Bibliometric analysis is a commonly used method for evaluating research trends [11]. To date, no bibliometric study has been conducted to systematically analyze the global literature on postoperative sleep over the past decade. Therefore, this study aimed to fill this gap by performing a bibliometric analysis of publications on postoperative sleep from 2014 to 2024. The specific research goals are as follows: (1) to characterize publication trends and identify key contributing countries, institutions, authors, and journals in this field; (2) to identify the core research hot spots and emerging topics; and (3) to provide a reference for researchers to grasp the global research status and formulate future research directions. The findings of this study will not only help academic researchers identify research gaps but also provide evidence-based insights for clinicians to optimize perioperative management strategies and improve patient recovery outcomes.

## Methods

### Ethical Considerations

This study did not include human or animal participants; therefore, ethics approval was not required for this study. For this type of study, formal consent was also not required.

### Identification of Papers on Postoperative Sleep

This study was a bibliometric analysis based on an accessible database. The WoS Core Collection database was systematically retrieved on May 6, 2025, to search for publications on postoperative sleep. The search terms were as follows: TS= (postoperative sleep disorder) OR TS= (postoperative sleep disturbance) OR TS= (postoperative insomnia) OR TS= (postoperative agrypnia) OR TS= (postoperative sleeplessness) OR TS= (postoperative hyposomnia) OR TS= (postoperative sleep quality). The time span was from 2014 to 2024.

Studies were excluded if they were non-English publications, publications on preoperative sleep disorders, non–surgery-related publications, publication types other than research papers and review papers, retracted publications, or publications on animal experiments. Two investigators independently searched and screened the records. The review process included two parts: (1) title and abstract screening (the 2 investigators independently reviewed the title and abstract of each retrieved publication) and (2) full-text verification (for publications with ambiguous descriptions in the titles and abstracts, the full texts were further retrieved and evaluated). If discrepancies arose, a third reviewer was consulted to resolve them through discussion or arbitration.

### Data Collection, Cleaning, and Analysis

Eligible publications were extracted. The extracted items included the title, authors, country or region, journal, affiliation, keywords, and publication year. The files were downloaded in ".txt" and ".xls" formats and then imported into VOSviewer (version 1.6.20) and Microsoft Excel (version 2021). All data extraction procedures were performed on the same day to avoid bias caused by database updates. In addition, the 2 investigators manually collected the specializations of the first authors.

Subsequently, we created thesaurus files to merge synonyms for countries and keywords. For example, "sleep disorders" and "insomnia" were replaced by "sleep disturbance." Thus, VOSviewer could automatically recognize and merge the duplicate terms. Quantitative analyses of the data, such as annual publication numbers and citation counts, were performed by Microsoft Excel.

VOSviewer was used to build a collaborative network of authors, countries, institutions, and journals to study collaborative linkages and the intensity of collaboration among different nodes, as well as to perform a clustering analysis of the frequently cocited keywords. For country, institution, and author collaboration networks, analysis type was coauthorship; the minimum number of papers per country, institution, and author were 5, 5, and 3, respectively; the normalization method was association strength; and the display style (meaning of the color) per country, institution, and author was by average publishing year, by average publishing years, and by cooperation strength, respectively. For the journal collaboration network, the analysis type was cocitation, the minimum number of papers per journal was 5, the normalization method was association strength, and the display style was by average publishing year. For the keyword co-occurrence network, the analysis type was co-occurrence, the minimum number of occurrences of a keyword was 10, the normalization method was association strength, and the display style was by average publishing year.

## Results

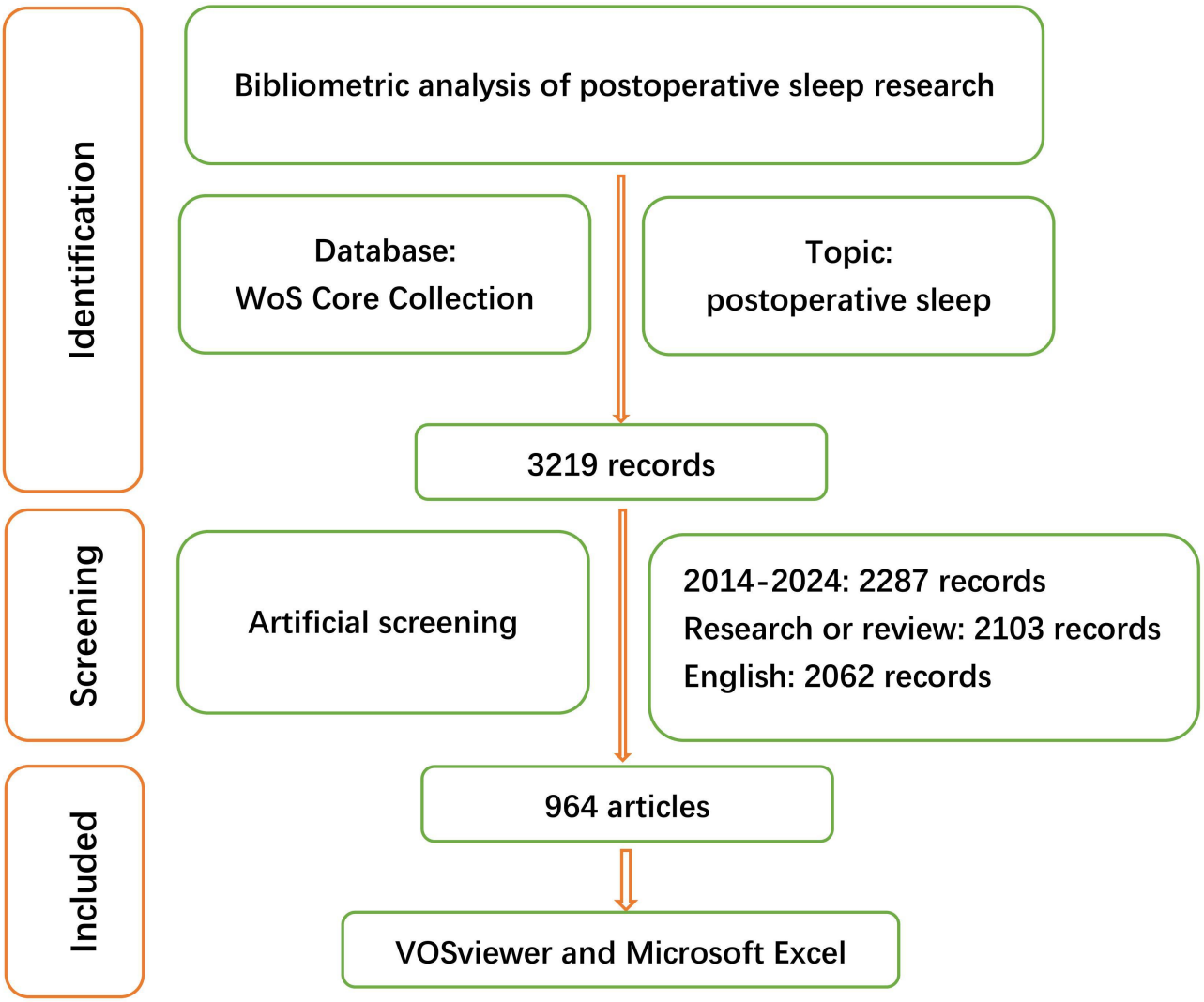
### Overview of the Literature Related to Postoperative Sleep

An initial search indicated that there were 3219 publications that may be eligible in the WoS Core Collection database. Among these publications, 2287 (71.0%) were published between 2014 and 2024. After removing non-English publications and other paper types, 2062 (64.1%) studies were reviewed by reading the titles and abstracts, and 1098 (53.2%) studies were excluded according to the exclusion criteria, with 964 (46.8%) studies finally included in this analysis. The details of this selection process are shown in Figure 1.



XSL•FO
RenderX

**Figure 1.** Flowchart demonstrating the selection process of this study. WoS: Web of Science.



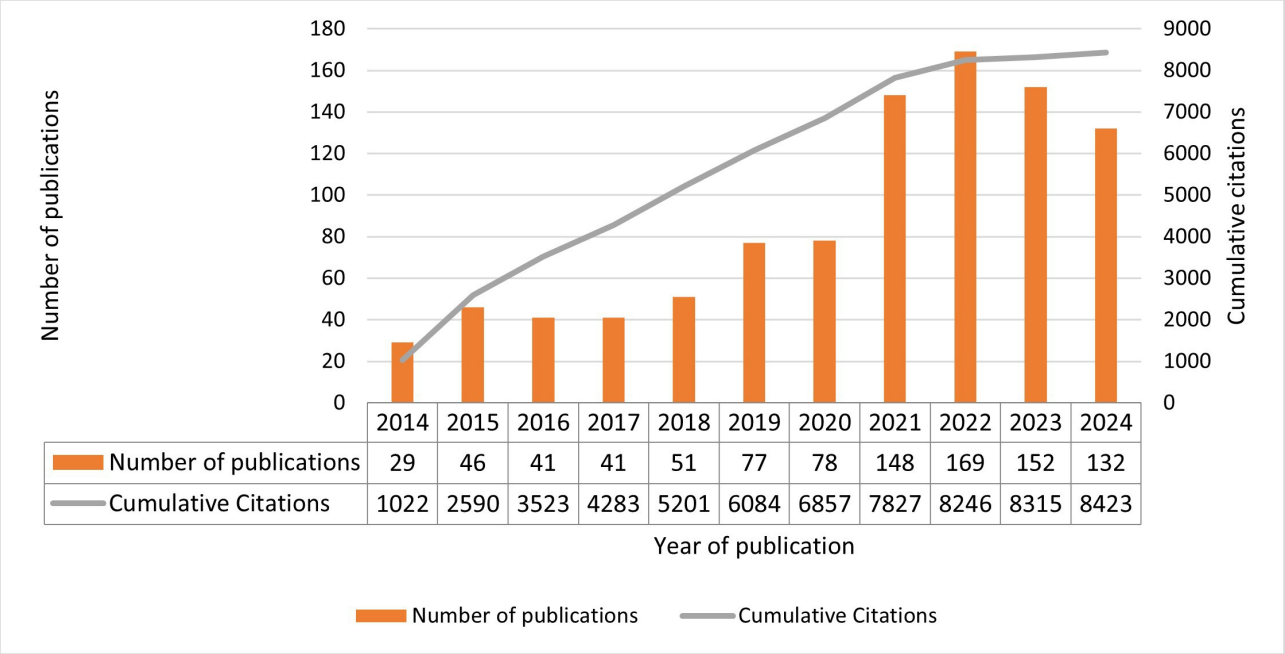

## Annual Trend of Publication Quantity

Postoperative sleep research has surged over the last decade. Figure 2 shows a substantial increase in annual publications, soaring from just 29 publications in 2014 to 169 publications in 2024. With a mean of 87.6 publications per year and a median of 77 publications per year, the field saw steady growth, punctuated by a significant surge in 2020. Research output has remained consistently high ever since, and these publications have already garnered an impressive 8423 citations.



XSL•FO
RenderX

**Figure 2.** Annual number of publications and cumulative citations.




| | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of publications | 29 | 46 | 41 | 41 | 51 | 77 | 78 | 148 | 169 | 152 | 132 |
| Cumulative Citations | 1022 | 2590 | 3523 | 4283 | 5201 | 6084 | 6857 | 7827 | 8246 | 8315 | 8423 |

## Contributions of Countries or Regions to Global Publications

In the last decade, 54 countries or regions contributed to the research on postoperative sleep. The top 10 countries or regions with the highest number of publications are listed in Table 1. China (n=402, 41.7%) made the highest contribution in terms of the number of publications, followed by the United States (n=201, 20.9%); Turkey (n=50, 5.2%); South Korea (n=48, 5.0%); Japan (n=35, 3.6%); Canada (n=26, 2.7%); Germany (n=25, 2.6%); Taiwan, China (n=23, 2.4%); Denmark (n=23, 2.4%); and Australia (n=19, 2.0%).

The top 10 countries or regions in terms of citation count were the United States (n=2587); China (n=2144); Canada (n=699); Denmark (n=515); South Korea (n=389); Taiwan, China (n=269); Germany (n=257); Australia (n=230); the United Kingdom (n=214); and France (n=210; Table 2). The United States had the most extensive collaboration links with other countries or regions, and the collaboration between the United States and China was strongest (Figure 3). In addition, papers from China, the United Kingdom, and Iran had younger publication ages (Figure 3).

**Table .** The top 10 countries or regions in terms of publications.

| Rank | Countries or regions | Publications, n |
|---|---|---|
| 1 | China | 402 |
| 2 | United States | 201 |
| 3 | Turkey | 50 |
| 4 | South Korea | 48 |
| 5 | Japan | 35 |
| 6 | Canada | 26 |
| 7 | Germany | 25 |
| 8 | Taiwan, China | 23 |
| 9 | Denmark | 23 |
| 10 | Australia | 19 |

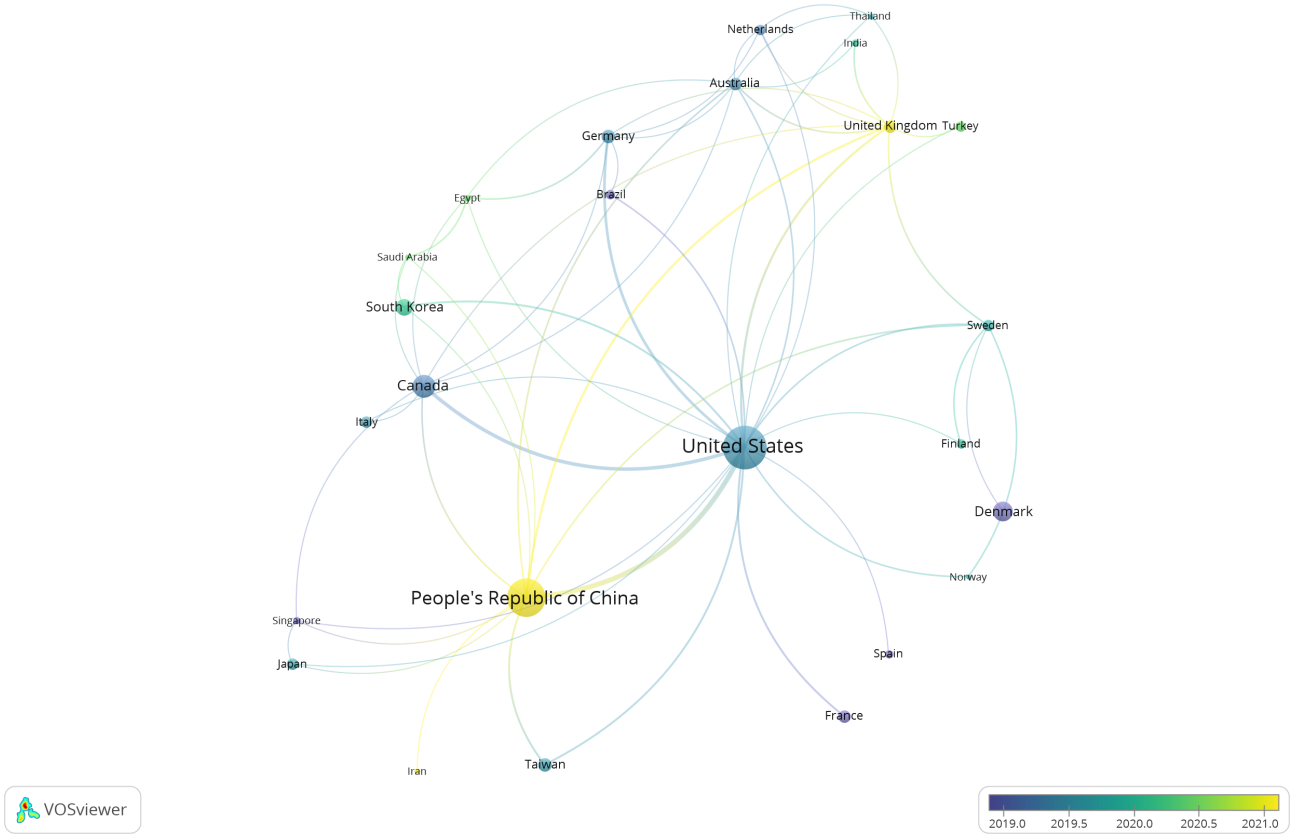


XSL•FO
RenderX

**Table .** The top 10 countries or regions in terms of citations.

| Rank | Countries or regions | Citations, n |
|---|---|---|
| 1 | United States | 2587 |
| 2 | China | 2144 |
| 3 | Canada | 699 |
| 4 | Denmark | 515 |
| 5 | South Korea | 389 |
| 6 | Taiwan, China | 269 |
| 7 | Germany | 257 |
| 8 | Australia | 230 |
| 9 | United Kingdom | 214 |
| 10 | France | 210 |

**Figure 3.** Links among the countries and regions. The nodes in the map have at least 5 papers.




## Contributions of the Authors

A total of 4677 authors published papers on postoperative sleep. Figure 4 shows 6 clusters, and the authors in each cluster had close citation relation. The top 10 authors in terms of publication number and citation count are listed in Tables 3 and 4. In terms of publication number, Zhu Junchao ranked first, followed by Song Bijia, Shi Qiuling, and Kehlet Henrik. In terms of citation number, Chung Frances ranked first, followed by Kehlet Henrik, Mace Jess C, and Smith Timothy L. In addition, the first authors mainly specialized in anesthesiology, surgery, internal medicine, and nursing.



XSL•FO
RenderX

**Figure 4.** Citation analysis of the authors. The nodes in the map have at least 3 papers.

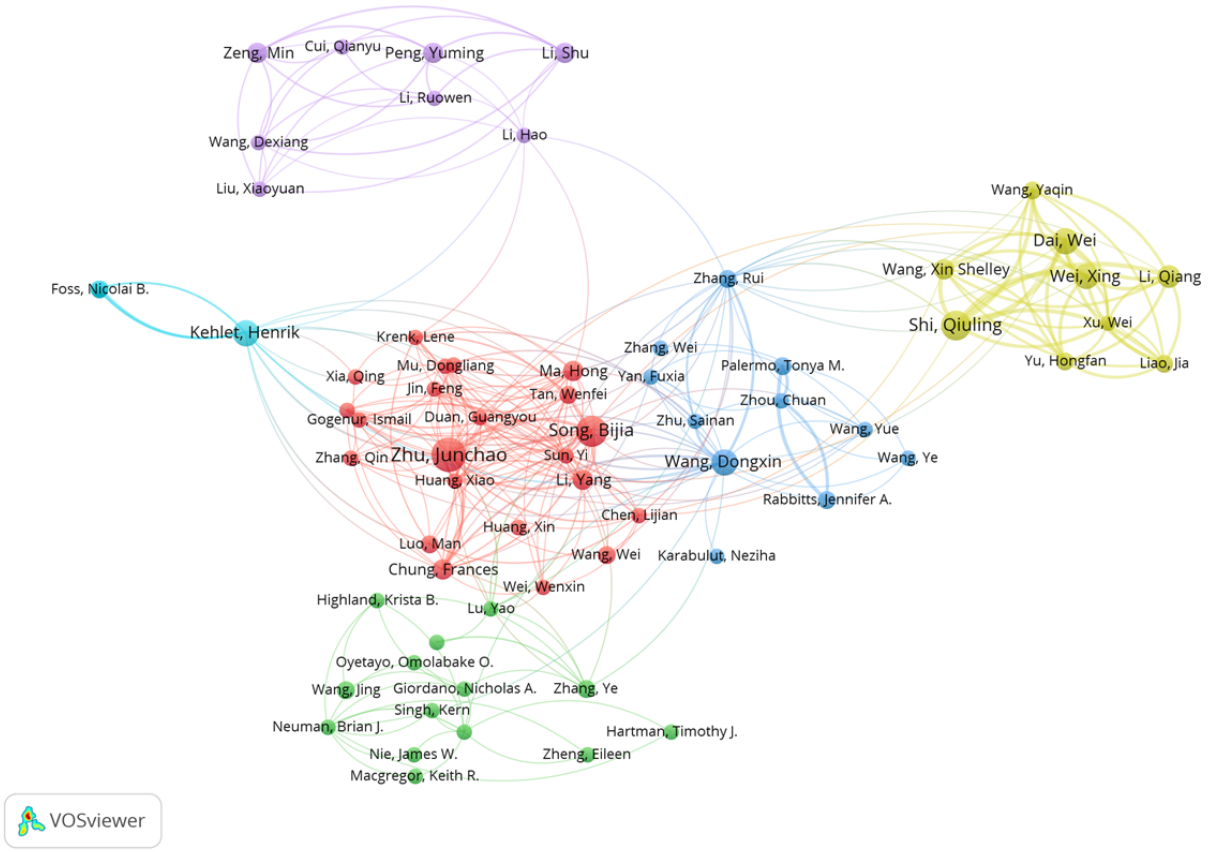



**Table .** The top 10 authors in terms of publications.

| Rank | Authors | Publications, n |
|---|---|---|
| 1 | Zhu Junchao | 17 |
| 2 | Song Bijia | 13 |
| 3 | Shi Qiuling | 11 |
| 4 | Kehlet Henrik | 10 |
| 5 | Wang Dongxin | 10 |
| 6 | Dai Wei | 9 |
| 7 | Wei Xing | 7 |
| 8 | Mace Jess C | 6 |
| 9 | Smith Timothy L | 6 |
| 10 | Li Qiang | 6 |



XSL•FO
RenderX

**Table .** The top 10 authors in terms of citations.

| Rank | Authors | Citations, n |
|---|---|---|
| 1 | Chung Frances | 248 |
| 2 | Kehlet Henrik | 227 |
| 3 | Mace Jess C | 219 |
| 4 | Smith Timothy L | 219 |
| 5 | Shi Qiuling | 161 |
| 6 | Wang Dongxin | 146 |
| 7 | Zhu Junchao | 122 |
| 8 | Wang Xin Shelley | 120 |
| 9 | Palermo Tonya M | 118 |
| 10 | Rabbitts Jennifer A | 118 |

## Contributions of the Affiliations

A total of 1153 scientific institutions contributed to these papers. The top 10 affiliations in terms of publication number are listed in Table 5. China Medical University (n=24) ranked first, followed by Capital Medical University (n=23), Sichuan University (n=19), Anhui Medical University (n=18), the University of Copenhagen (n=18), and Peking University (n=18). In terms of citation number, the University of Copenhagen ranked first (n=488), followed by the University of Toronto (n=309), Peking University (n=232), and Oregon Health and Science University (n=224). Universities such as the University of Toronto and Stanford University were active in the 2010s, whereas universities such as Tongji University and Duke University have devoted themselves to this area more recently (Figure 5).

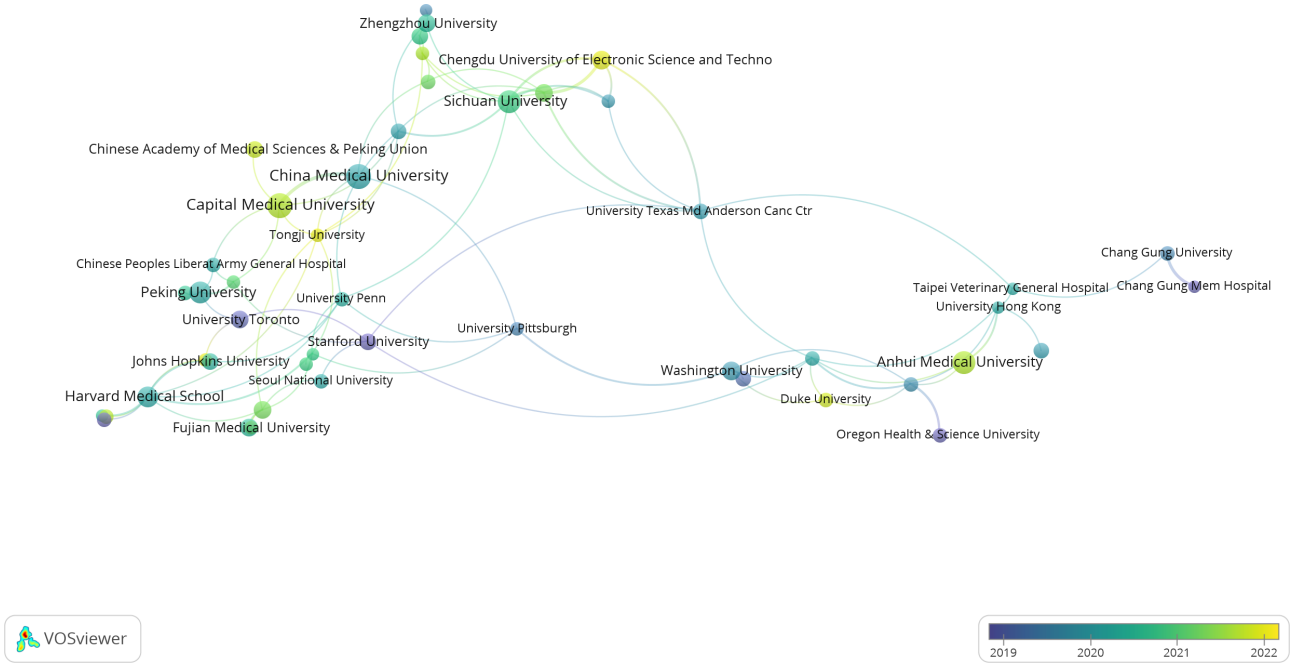
**Table .** The top 10 institutions with the most number of publications.

| Rank | Institutions | Publications, n | Citations, n | Countries or regions |
|---|---|---|---|---|
| 1 | China Medical University | 24 | 183 | China |
| 2 | Capital Medical University | 23 | 105 | China |
| 3 | Sichuan University | 19 | 148 | China |
| 4 | Anhui Medical University | 18 | 98 | China |
| 5 | University of Copenhagen | 18 | 488 | Denmark |
| 6 | Peking University | 18 | 232 | China |
| 7 | Harvard Medical School | 14 | 169 | United States |
| 8 | University of Electronic Science and Technology of China | 11 | 74 | China |
| 9 | Washington University | 11 | 161 | United States |
| 10 | Chongqing Medical University | 10 | 104 | China |
| 11 | Shanghai Jiao Tong University | 10 | 69 | China |
| 12 | Zhengzhou University | 10 | 97 | China |
| 13 | University of Toronto | 10 | 309 | Canada |
| 14 | Fujian Medical University | 10 | 121 | China |



XSL•FO
RenderX

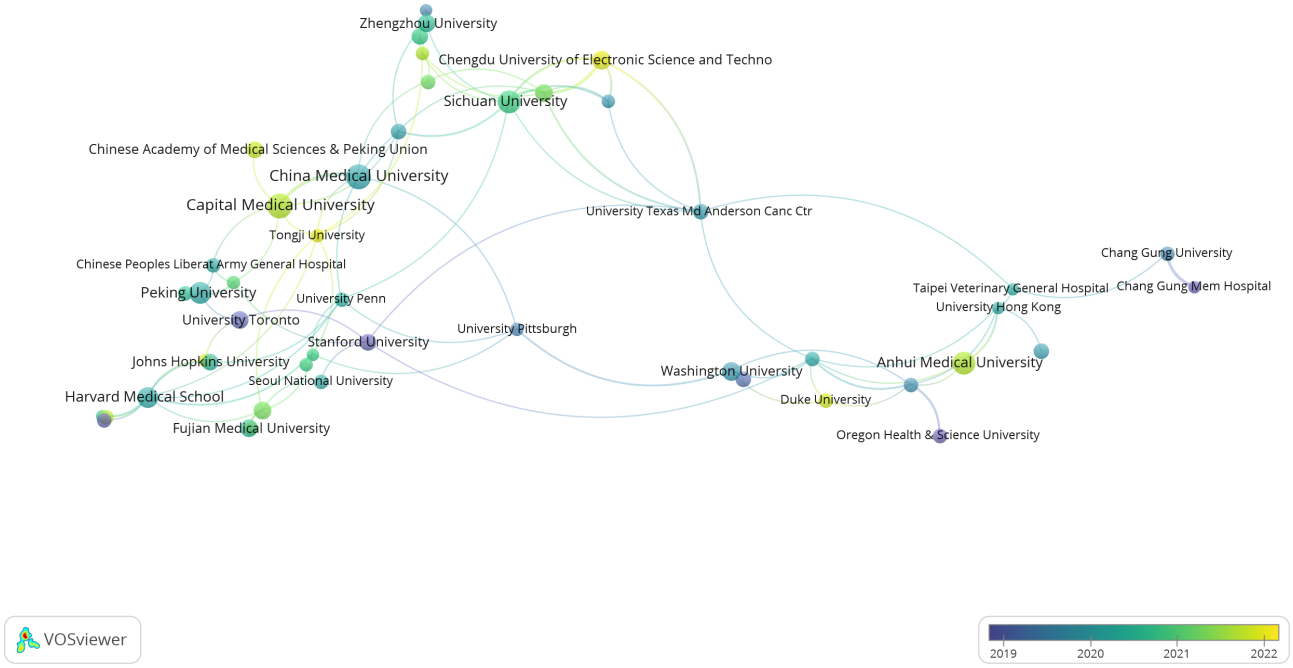
**Figure 5.** Citation relationships among institutions. The minimum number of papers per institution was set at 5.




## Distribution of the Journals

The papers on postoperative sleep in the last 10 years were documented by 392 journals indexed in the Scientific Citation Index. *BMC Anesthesiology*, with 19 publications, ranked first, followed by *Nature and Science of Sleep* (n=16) and *International Journal of Clinical and Experimental Medicine* (n=16). Papers published in the *British Journal of Anaesthesia*, *Journal of Clinical Anesthesia*, *Anesthesia and Analgesia*, *Journal of Pain*, and *Journal of Sleep Medicine* had higher citation counts on average (Figure 6).

**Figure 6.** Citation relationships among journals. The minimum number of papers per journal was set at 5.




## Analysis of Keywords

The keyword co-occurrence analysis (Figure 7) revealed the primary research themes and focal points within the domain of postoperative sleep disturbances. The central and most prominent keywords were postoperative pain, sleep disturbance, and quality of life, indicating that these topics are core concerns driving research. Patient-reported outcomes also feature significantly, highlighting the importance of subjective patient experiences in this field.

Research methodologies are notably represented by specific assessment tools, including actigraphy (objective sleep measurement), polysomnography (the gold standard sleep study), and the Pittsburgh Sleep Quality Index (a common subjective sleep quality questionnaire). Key clinical contexts strongly associated with postoperative sleep issues included



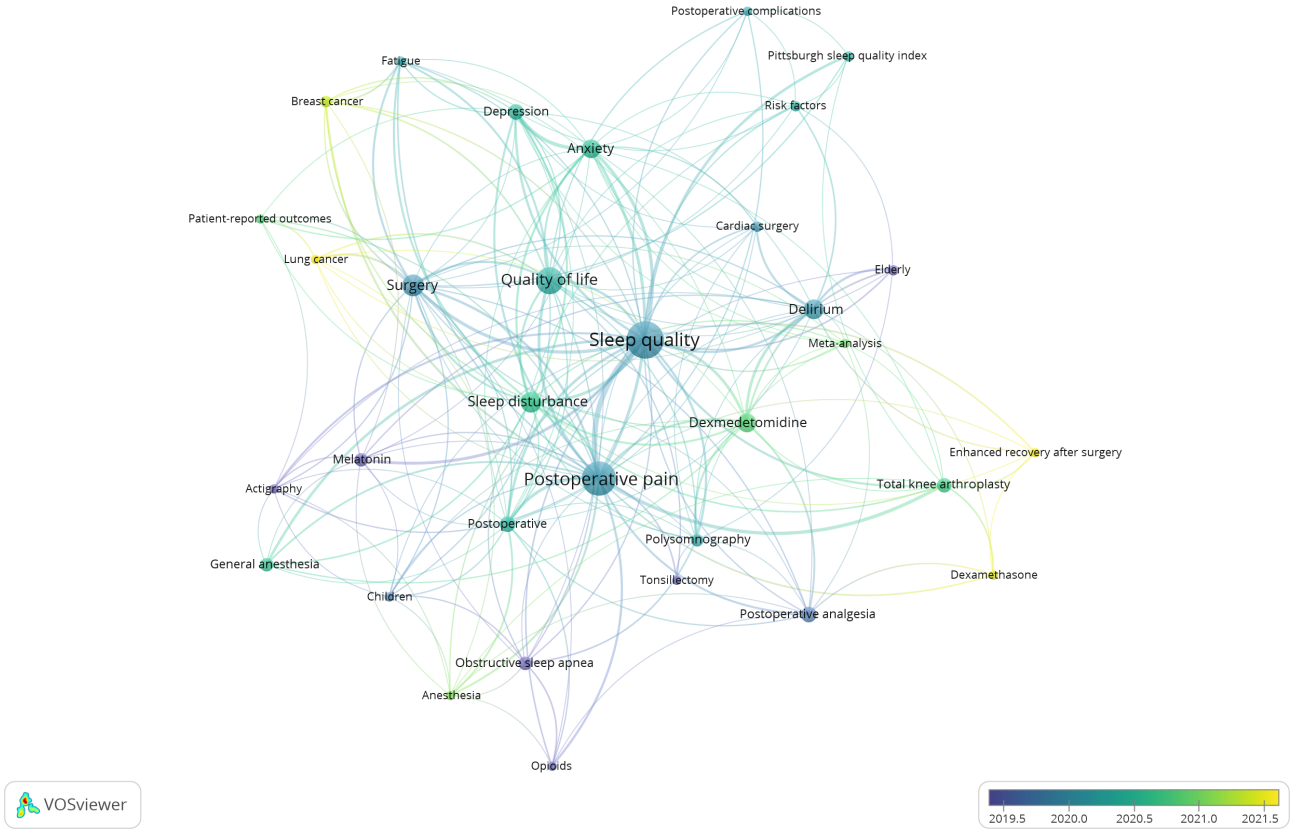
XSL•FO
RenderX

breast cancer, lung cancer, cardiac surgery, total knee arthroplasty, and tonsillectomy.

Pharmacological interventions such as dexmedetomidine and melatonin were also frequently identified. The concept of enhanced recovery after surgery (ERAS) appears frequently, suggesting that its protocols are being examined for their impact on preoperative sleep. Other significant keywords include "fatigue" and "delirium" (a serious postoperative complication potentially linked to sleep disturbance).

The temporal markers (2019.5 to 2021.5) indicate that this topic has received particular attention in recent years. "ERAS," "lung cancer," and "dexamethasone" emerge as more recent keywords in postoperative sleep research.

**Figure 7.** Co-occurrence analysis of author-provided keywords. The minimum occurrence threshold for inclusion in the map was set at 10.




## Discussion

### Principal Findings

This bibliometric analysis focused on papers related to postoperative sleep over the past 10 years and provided an overview of publication trends, the distribution of publications, and hot research topics.

On the basis of the data collected in the WoS Core Collection, 964 papers on postoperative sleep were published during the past decade. With an increasing trend, the number of publications remained high. In total, 4677 authors from 1153 institutions across 54 countries or regions worldwide have published their work, and the number of researchers continues to grow. This trend and the distribution of publications suggest that postoperative sleep is an important focus in clinical practice and a prominent topic in the scientific literature, consistent with previous findings that postoperative sleep disorders can affect recovery [10,12,13].

China and the United States contributed the most to postoperative sleep research. They had the highest number of publications and citations and demonstrated the strongest international collaborations. However, fresh institutions in terms of average publication years mainly come from China, which indicates the expanding participation in this field. It is interesting that the top 10 authors and institutions in terms of publication number and citations were all from high-income countries or regions. This finding is not surprising but reasonable. A previous study was conducted to uncover the relationship between the development of anesthesia and gross domestic product from 2009 to 2018, and the results indicated a strong positive correlation [14].

Zhu Junchao is the most productive author in this field. The author has long been devoted to investigating the relationships between organ protection and cognitive impairment in the field of ischemia-reperfusion injury, postoperative outcomes, and sleep quality [15-17]. One of the related studies found that morning operations required a higher dose of anesthetics than evening operations and that the degree of postoperative sleep disorders was greater when the operation was performed in the evening than in the morning [17]. Meanwhile, the results prompted that patients with hyperpathia and somnipathy might benefit from operations performed in the morning [17,18]. Chung Frances, with 248 citations and only 5 publications, is the most cited author in the field of postoperative sleep research from 2014 to 2024. However, he and his research team have



XSL•FO
RenderX

been engaged in postoperative sleep disorders for more than 10 years [13,19,20]. His paper published in *Anesthesiology* in 2014 revealed that the postoperative sleep structure was disturbed in both patients with obstructive sleep apnea and patients with nonobstructive sleep apnea, and they further analyzed disturbances in sleep structure [19]. His systematic review and meta-analysis also revealed the prevalence and risk factors of sleep disturbances in surgical patients [20]. These aforementioned authors are well respected as they greatly contribute to the prevention of postoperative sleep disturbance.

It was reported that the incidence of sleep disturbance was high (54%) during the postoperative period [21]. Determining the prevalence and risk factors related to postoperative sleep disturbance would be beneficial for risk stratification and perioperative care planning. On the basis of the keywords map, sleep quality was closely related to the postoperative quality of life, postoperative pain, delirium, anxiety, depression, dexmedetomidine use, and postoperative complications, while ERAS, lung cancer, breast cancer, and dexamethasone have received much attention recently. Postoperative pain is regarded as a primary risk factor for postoperative sleep disorders, and sleep quality can, in turn, aggravate pain [22-25].

The term “enhanced recovery after surgery” was coined by Henrik Kehlet, one of the most influential researchers in the field of postoperative sleep. Postoperative pain, a key component of ERAS protocols, and postoperative sleep have received considerable attention. Opioids, nonopioids, regional and neuraxial anesthesia, and multimodal analgesia based on the aforementioned techniques have been widely used in ERAS procedures, contributing to improved postoperative pain control [26,27]. Nevertheless, the incidence of postoperative sleep disorders is still high even though postoperative pain is well controlled [28,29]. Future studies are advised to build a systematic strategy for postoperative management.

To our knowledge, this is the first bibliometric analysis based on the WoS Core Collection to assess the studies on postoperative sleep over the past 10 years. This study included a large number of papers (n=964) on postoperative sleep to help identify the research status and hot spots. To improve the quality of the enrolled studies, 2 independent investigators conducted an artificial screening, and disagreements were properly settled. Moreover, duplicate terms were removed as far as possible by creating a thesaurus file.

The findings of this bibliometric analysis provide several practical implications for clinicians engaged in perioperative patient management. First, postoperative pain was identified as a core correlate of sleep disturbance in this study. Practitioners should integrate sleep quality assessment and interventions into ERAS pathways. They can combine multimodal analgesia with sleep-promoting strategies, such as melatonin administration or environmental optimization, to improve both pain control and sleep quality in surgical patients. Second, keyword analysis highlighted lung cancer and breast cancer surgeries as emerging research hot spots. Practitioners should pay special attention to sleep disorders in patients undergoing these procedures.

Third, the analysis shows that anesthesiologists, surgeons, and nurses are the main contributors to this field. Practitioners should establish interdisciplinary teams to address postoperative sleep issues. For example, anesthesiologists can optimize anesthetic regimens to reduce circadian rhythm disruption, while nurses can implement nonpharmacological sleep interventions during inpatient care. In addition, the identification of emerging topics such as ERAS provides direction for future clinical research. Practitioners can design prospective studies to verify the effects of postoperative sleep within ERAS protocols.

However, this study inevitably has some limitations. It was impossible to enroll all publications in this field, and we restricted the paper type to research and review papers. Therefore, the results of this study represent the general situation of postoperative sleep research but cannot precisely reflect all details. We did not use processed parameters such as the Hirsch index to define the top authors and institutions because the primary function of bibliometric analysis is to provide a macrolevel view of a research field. Citation indicators may be influenced by the time of publication, meaning that recently published studies may have been underestimated. The number of citations increased dynamically over time. Thus, bibliometric analyses in this field must be updated in the future.

## Conclusions

This study presents the current research status and possible research directions through a visual bibliometric analysis of research on postoperative sleep over the past 10 years. There has been a large amount of literature on postoperative sleep over the past 10 years. Authors and organizations from China are leading contributors, followed by those from the United States. Anesthesiology is a critical discipline in this field. Postoperative pain is closely related to postoperative sleep and has become a hot topic. Recent studies have mainly focused on lung and breast cancer surgeries, and ERAS has become an emerging keyword.

## Acknowledgments

The authors thank all individuals who participated in and supported this study. They declare that no generative artificial intelligence was used in any part of the manuscript preparation.

## Funding

This study did not receive financial support from any organization or individual.



XSL•FO
RenderX

## Conflicts of Interest

None declared.

## References


1. Stenvers DJ, Scheer F, Schrauwen P, la Fleur SE, Kalsbeek A. Circadian clocks and insulin resistance. Nat Rev Endocrinol 2019 Feb;15(2):75-89. [doi: 10.1038/s41574-018-0122-1] [Medline: 30531917]
2. Laver KE, Spargo C, Saggese A, et al. Sleep disturbance and disorders within adult inpatient rehabilitation settings: a systematic review to identify both the prevalence of disorders and the efficacy of existing interventions. J Am Med Dir Assoc 2020 Dec;21(12):1824-1832.e2. [doi: 10.1016/j.jamda.2020.03.002] [Medline: 32312680]
3. Wang J, Lu FF, Ge MM, et al. Transcutaneous electrical acupoint stimulation improves postoperative sleep quality in patients undergoing laparoscopic gastrointestinal tumor surgery: a prospective, randomized controlled trial. Pain Ther 2023 Jun;12(3):707-722. [doi: 10.1007/s40122-023-00493-2] [Medline: 36928500]
4. Poulsen RC, Warman GR, Sleigh J, Ludin NM, Cheeseman JF. How does general anaesthesia affect the circadian clock? Sleep Med Rev 2018 Feb;37:35-44. [doi: 10.1016/j.smrv.2016.12.002] [Medline: 28162920]
5. van Zuylen ML, Meewisse AJG, Ten Hoope W, et al. Effects of surgery and general anaesthesia on sleep-wake timing: CLOCKS observational study. Anaesthesia 2022 Jan;77(1):73-81. [doi: 10.1111/anae.15564] [Medline: 34418064]
6. McCann ME, Soriano SG. Does general anesthesia affect neurodevelopment in infants and children? BMJ 2019 Dec 9;367:l6459. [doi: 10.1136/bmj.l6459] [Medline: 31818811]
7. Srinivasan M, Walker C. Circadian clock, glucocorticoids and NF-κB signaling in neuroinflammation- implicating glucocorticoid induced leucine zipper as a molecular link. ASN Neuro 2022;14:17590914221120190. [doi: 10.1177/17590914221120190] [Medline: 36317290]
8. Su X, Wang DX. Improve postoperative sleep: what can we do? Curr Opin Anaesthesiol 2018 Feb;31(1):83-88. [doi: 10.1097/ACO.0000000000000538] [Medline: 29120927]
9. Yılmaz M, Karabulut N. How can we improve the comfort level and sleep quality after surgery? J Perianesth Nurs 2022 Feb;37(1):100-104. [doi: 10.1016/j.jopan.2021.05.008] [Medline: 34836766]
10. Pitaro NL, Herrera MM, Alasadi H, et al. Sleep disturbance trends in the short-term postoperative period for patients undergoing total joint arthroplasty. J Am Acad Orthop Surg 2023 Oct 1;31(19):e859-e867. [doi: 10.5435/JAAOS-D-23-00059] [Medline: 37523691]
11. Ling F, Qi W, Li X, et al. Bibliometric analysis of acupuncture therapy for cancer pain over the past 10 years. J Pain Res 2023 Mar 20;16:985-1003. [doi: 10.2147/JPR.S395421] [Medline: 36968762]
12. Many BT, Rizeq YK, Kwon S, et al. Quantifying postoperative sleep loss associated with increased pain in children undergoing a modified Nuss operation. J Pediatr Surg 2020 Sep;55(9):1846-1849. [doi: 10.1016/j.jpedsurg.2019.12.003] [Medline: 31982091]
13. Mokhlesi B, Hovda MD, Vekhter B, Arora VM, Chung F, Meltzer DO. Sleep-disordered breathing and postoperative outcomes after elective surgery: analysis of the nationwide inpatient sample. Chest 2013 Sep;144(3):903-914. [doi: 10.1378/chest.12-2905] [Medline: 23538745]
14. Doğan G, Karaca O. A bibliometric analysis of the field of anesthesia during 2009-2018 [Article in Portuguese]. Braz J Anesthesiol 2020;70(2):140-152. [doi: 10.1016/j.bjan.2020.02.003] [Medline: 32499046]
15. Gu X, Zhang Y, Wei W, Zhu J. Effects of preoperative anxiety on postoperative outcomes and sleep quality in patients undergoing laparoscopic gynecological surgery. J Clin Med 2023 Feb 24;12(5):1835. [doi: 10.3390/jcm12051835] [Medline: 36902622]
16. Song B, Zhu J. A narrative review of the impact of cerebellar dysfunction and sleep disturbances after general anesthesia in patients with Alzheimer's disease. Ann Palliat Med 2021 Jun;10(6):6919-6925. [doi: 10.21037/apm-20-2597] [Medline: 33977757]
17. Song B, Li Y, Teng X, Li X, Yang Y, Zhu J. Comparison of morning and evening operation under general anesthesia on intraoperative anesthetic requirement, postoperative sleep quality, and pain: a randomized controlled trial. Nat Sci Sleep 2020;12:467-475. [doi: 10.2147/NSS.S257896] [Medline: 32765143]
18. Song B, Li Y, Teng X, Li X, Yang Y, Zhu J. The effect of intraoperative use of dexmedetomidine during the daytime operation vs the nighttime operation on postoperative sleep quality and pain under general anesthesia. Nat Sci Sleep 2019 Oct 3;11:207-215. [doi: 10.2147/NSS.S225041] [Medline: 31686933]
19. Chung F, Liao P, Yegneswaran B, Shapiro CM, Kang W. Postoperative changes in sleep-disordered breathing and sleep architecture in patients with obstructive sleep apnea. Anesthesiology 2014 Feb;120(2):287-298. [doi: 10.1097/ALN.0000000000000040] [Medline: 24158049]
20. Butris N, Tang E, Pivetta B, et al. The prevalence and risk factors of sleep disturbances in surgical patients: a systematic review and meta-analysis. Sleep Med Rev 2023 Jun;69:101786. [doi: 10.1016/j.smrv.2023.101786] [Medline: 37121133]
21. Hamai S, Harada S, Tsushima H, et al. Interaction between functional capability and sleep quality at midterm after total knee arthroplasty: a Japanese retrospective cohort study. Sci Rep 2023 Oct 26;13(1):18373. [doi: 10.1038/s41598-023-45603-4] [Medline: 37884595]




XSL•FO
RenderX

22. Yang MMH, Riva-Cambrin J, Cunningham J, Casha S. Impact of preoperative insomnia on poor postoperative pain control after elective spine surgery and the modified Calgary Postoperative Pain After Spine Surgery (MCAPPS) score. N Am Spine Soc J 2023 Aug 6;16:100261. [doi: 10.1016/j.xnsj.2023.100261] [Medline: 37753373]
23. Zhang Z, Wang H, Wang Y, Luo Q, Yuan S, Yan F. Risk of postoperative hyperalgesia in adult patients with preoperative poor sleep quality undergoing open-heart valve surgery. J Pain Res 2020 Oct 13;13:2553-2560. [doi: 10.2147/JPR.S272667] [Medline: 33116797]
24. O'Gara BP, Gao L, Marcantonio ER, Subramaniam B. Sleep, pain, and cognition: modifiable targets for optimal perioperative brain health. Anesthesiology 2021 Dec 1;135(6):1132-1152. [doi: 10.1097/ALN.0000000000004046] [Medline: 34731233]
25. Liu D, Li X, Nie X, et al. Artificial intelligent patient-controlled intravenous analgesia improves the outcomes of older patients with laparoscopic radical resection for colorectal cancer. Eur Geriatr Med 2023 Dec;14(6):1403-1410. [doi: 10.1007/s41999-023-00873-z] [Medline: 37847474]
26. Simpson JC, Bao X, Agarwala A. Pain management in enhanced recovery after Surgery (ERAS) protocols. Clin Colon Rectal Surg 2019 Mar;32(2):121-128. [doi: 10.1055/s-0038-1676477] [Medline: 30833861]
27. Joshi GP, Kehlet H. Postoperative pain management in the era of ERAS: an overview. Best Pract Res Clin Anaesthesiol 2019 Sep;33(3):259-267. [doi: 10.1016/j.bpa.2019.07.016] [Medline: 31785712]
28. Qiu D, Wang XM, Yang JJ, et al. Effect of intraoperative esketamine infusion on postoperative sleep disturbance after gynecological laparoscopy: a randomized clinical trial. JAMA Netw Open 2022 Dec 1;5(12):e2244514. [doi: 10.1001/jamanetworkopen.2022.44514] [Medline: 36454569]
29. Huang X, Wang J, Zhang J, Kang Y, Sandeep B, Yang J. Ultrasound-guided erector spinae plane block improves analgesia after laparoscopic hepatectomy: a randomised controlled trial. Br J Anaesth 2022 Sep;129(3):445-453. [doi: 10.1016/j.bja.2022.05.013] [Medline: 35803754]


## Abbreviations

**ERAS:** enhanced recovery after surgery
**WoS:** Web of Science


*Edited by M Balcarras; submitted 18.Oct.2025; peer-reviewed by A Al-Asadi, K Pang; revised version received 01.Feb.2026; accepted 26.Feb.2026; published 26.Mar.2026.*

*Please cite as:*
*Wei X, Fu L, Jiang W, Zhang X, Zhou R*
*Mapping the Global Surge in Postoperative Sleep Research From 2014 to 2024: Bibliometric Analysis*
*Interact J Med Res 2026;15:e86086*
*URL: https://www.i-jmr.org/2026/1/e86086*
*doi:10.2196/86086*


© Xin Wei, Lijuan Fu, Wencai Jiang, Xianjie Zhang, Rui Zhou. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 26.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Sleep Disturbance and Its Association With Purchasing Behavior of COVID-19 Medicine Among the Public After the Adjustment of Zero-COVID Policy in China: Results From a Web-Based Survey Study

Yu Huang[1], MSc; Xiang Zhao[1], MSc; Lei Wang[1], MSc; Qiaohong Lv[1], MSc; Qingqing Wu[1], MSc; Shuiyang Xu[1], MPH; Xuehai Zhang[1], MSc; Shuxian Wu[2], MSc

[1]Department of Health Education, Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China

[2]Central Office, Zhejiang Provincial Center for Disease Control and Prevention, 630 Xincheng Road, Binjiang District, Hangzhou, Zhejiang, China

**Corresponding Author:**
Shuxian Wu, MSc
Central Office, Zhejiang Provincial Center for Disease Control and Prevention, 630 Xincheng Road, Binjiang District, Hangzhou, Zhejiang, China

## Abstract

**Background:** In December 2022, in light of the weakened pathogenicity of the new variants and other scientific considerations, China optimized its zero-COVID policy. As the situation evolved, the virus spread more widely across the country.

**Objective:** This study aims to explore the public's sleep status and its association with purchasing behavior of COVID-19 medicine after the adjustment of zero-COVID policy in China.

**Methods:** A cross-sectional, internet-based survey among residents aged 18 - 69 years was conducted in Zhejiang province, China, from December 16 to 30, 2022, to collect data on sociodemographic characteristics, COVID-19 drug purchasing behavior, sleep disturbance levels, etc. Chi-square tests, univariate analyses, and multivariate analyses were used to explore the associations among these factors.

**Results:** Out of 38,480 participants, 20,803 (54.1%) reported sleep disruption after China's COVID-19 response policy adjustment. The degree of impact varied, with 10,964 (52.70%) reporting "slight," 3105 (14.93%) "moderate," 3493 (16.79%) "significant," and 3241 (15.58%) "very significant" impacts. Only 20.90% (782/3742) of those who deemed purchasing unnecessary had sleep disruptions, compared to 45.19% (6214/13,752) of those who acquired medications and 65.79% (13,807/20,986) of those who tried but failed to obtain them. Sleep disturbance levels were significantly associated with sociodemographic factors like age, education levels, occupation, marital status, and presence of family members diagnosed with COVID-19 ($P$<.05). By age, sleep disturbance proportions differed notably: 36.32% (409/1126) for those under 20 years, 54.81% (19,714/35,970) for the 20 to 60 age group, and 49.13% (680/1384) for individuals over 60 years. For education level, the proportions were 57.44% (517/900, primary school), 54.34% (3928 /7229, junior high school), 54.27% (3808/7017, senior high school), 53.99% (11,974/22,180, junior college/undergraduate), and 49.91% (576/1154, master's degree), showing a clear downward trend as education level increased. By occupation, farmers had the highest rate (855/1447, 59.09%), followed by business/service industry workers and stay-at-home/unemployed individuals (13,925/24,750, 56.26%) and government staff (4161/7712, 53.95%), while 1242 out of 3049 (40.73%) health workers and 620 out of 1522 (40.74%) students had lower rates. Married participants had a 55.21% (17,143/31,053) sleep disturbance rate, and those with COVID-positive family members had the highest rate (2023/2873, 70.41%). Multivariate logistic regression, adjusting for these sociodemographic factors, showed that compared to those who thought purchasing COVID-19 medications was unnecessary, those who acquired medications were 3.11 times (adjusted odds ratio 3.11, 95% CI 2.85 - 3.39) more likely, and those who tried but couldn't get medications were 7.11 times (adjusted odds ratio 7.11, 95% CI 6.53 - 7.74) more likely to experience sleep disturbance.

**Conclusions:** The adjustment of China's zero-COVID policy affected the sleep health of the public, which was closely linked to drug-purchasing status, especially among the older people, those with lower education levels, and those with family members diagnosed with COVID-19. It highlights the need to develop and deploy interventions aimed at promoting better sleep health in times of crisis.

***(Interact J Med Res 2026;15:e79903)*** doi:10.2196/79903

**KEYWORDS**

adjustment of zero-COVID policy; China; COVID-19; drug purchasing; drug shortage; sleep disturbance



XSL•FO
RenderX

## Introduction

In December 2019, an unprecedented pneumonia outbreak of unknown origin emerged in Wuhan, Hubei province, China [1,2]. Subsequently identified as caused by the novel coronavirus (COVID-19), this highly contagious pathogen rapidly spread beyond China's borders within just 30 days [3-6]. Characterized by potentially severe symptoms such as acute respiratory failure and sepsis [7], COVID-19 posed a significant threat to global public health, prompting the World Health Organization (WHO) to declare it a public health emergency of international concern on January 30, 2020 [8].

To contain the virus, China implemented the dynamic zero-COVID policy, which included strict measures like national lockdowns, school closures, travel restrictions, and the suspension of numerous economic activities [9-13]. However, as the virulence of new virus variants decreased and COVID-19 vaccines became more prevalent, China adjusted its policy on December 7, 2022. The mandatory large-scale nucleic acid testing requirements, the centralized isolation for asymptomatic cases and mild symptomatic patients, and the mobility restrictions were removed [14,15]. This suddenly loosened policy led to a change in the epidemic situation, with a wider spread of the virus across the country.

The COVID-19 pandemic has had a profound impact on people's lives, particularly on their mental health and sleep quality [16-20]. A large-scale global study (combining 250 studies with nearly 500,000 participants from 49 countries) found that about 40% of people had sleep disturbances during the COVID-19 pandemic (95% CI 37.56%-43.48%). Among specific populations, the prevalence of sleep problems varied, with 52.39% (95% CI 41.69%-62.88%) among COVID-19 patients, 45.96% (95% CI 36.90%-55.30%) among children and adolescents, 42.47% (95% CI 37.95%-47.12%) among health care workers, 41.50% (95% CI 32.98%-50.56%) among special populations with health care needs, 41.16% (28.76%-54.79%) among university students, and 36.73% (32.32%-41.38%) among the general population [21]. Research in China shows that the prevalence of insomnia during the early and late stages of the pandemic was 37.0% (95% CI 34.1% - 39.9%) and 41.8% (95% CI 33.6% - 50.0%) respectively. Notably, health care workers, COVID-19 patients, those with chronic diseases, and people with mental disorders are more likely to experience insomnia than the general population [22].

Sleep is a fundamental physiological process that is essential for maintaining physical and mental health [19,23]. It plays a crucial role in various bodily functions, including homeostasis, immune system regulation, cognitive function, neural plasticity, memory consolidation, energy maintenance, macromolecular biosynthesis, and metabolism regulation [24,25]. Conversely, sleep deprivation can lead to a range of adverse health outcomes, such as impaired physical and mental performance, increased risk of depression, stroke, chronic inflammation, cancers, and weakened immune defense, which may be particularly relevant during the COVID-19 pandemic [26,27].

During the pandemic, multiple factors influenced people's sleep quality. The spread of news and rumors through social media, television, and daily interactions about COVID-19's severity and fatality numbers fueled anxiety and fear. Additionally, lack of in-person social connections, changes in eating patterns, less outdoor activities which lead to limited exposure to sunlight, and consequently extended exposure to blue and bright light from screens may lead to disturbed circadian rhythms [21,28-32]. Above all, the suddenly loosened policy and the increased risk of infection may lead some individuals to panic-purchase medical resources—especially COVID-19 medicines (including cough medicine, nasal congestion/snot relievers, fever reducers, pain relievers, and other symptom-alleviating pharmaceuticals for COVID-19)—to ensure timely administration when needed [33,34]. The purchasing behavior towards COVID-19 medicines reflected people's perception of the disease and their self-protection strategies. However, it may also leave those in need unable to obtain them. This behavior was not only driven by the actual need for medications but also by psychological factors, such as the desire for a sense of security in the face of an uncertain health situation. Understanding the relationship between sleep status and these factors after the policy adjustment is of great significance for public health.

However, there is a lack of in-depth research on how the adjustment of China's zero-COVID policy affects the public's sleep status and its connection with the purchasing behavior of COVID-19 medicine. This study aims to fill this gap by conducting a cross-sectional, internet-based survey to assess the public's sleep status and explore its association with the purchasing behavior of COVID-19 medicine after the policy adjustment. The findings of this research can provide valuable insights for public health interventions and future pandemic-related policies.

## Methods

### Questionnaire

The questionnaire was formulated by the Zhejiang Provincial Center for Disease Control and Prevention based on pandemic-specific context, literature review, and expert consultation. To ensure its reliability, a pilot study was first conducted. Following the feedback from the pilot study and through two rounds of collaboration based on the Delphi method, the questionnaire was revised. During this process, subject matter experts evaluated its face validity, and a literature review was carried out to assess its content validity.

The questionnaire collected sociodemographic data: purchasing behavior of COVID-19 medications; specific types of COVID-19 drugs that respondents had purchased; channels through which they bought these medications; and degrees of their sleep disturbances. Sociodemographic data includes age, gender, education level, occupation, marital status, and diagnosis of COVID-19 among family members. The specific questions regarding the purchasing behavior of COVID-19 medications and the degrees of their sleep disturbances were as follows: (1) "Have you purchased COVID-19 drugs?" (2) "Which of the following COVID-19 drugs have you purchased?" (3) "Through



XSL•FO
RenderX

what channel did you successfully purchase these medicines?" (4) "In the past week, has the adjustment of COVID-19 policies affected your sleep?" No item randomization or adaptive questioning was used, and the 18-item questionnaire was distributed across 3 screens. Participants could review and revise answers via a "Back" button on each screen before submission. IP address tracking was used to prevent multiple entries from the same individual, and no two entries from the same IP address were allowed within 72 hours.

To maintain the quality of the survey data, a quality control question was incorporated into the questionnaire to identify survey respondents who did not answer the questions seriously. The question was, "This is a quality control question, please choose B." All the survey subjects who chose other options were not included in the final data analysis. In addition, server-side completeness checks were implemented upon submission, incomplete responses were rejected, and participants were prompted to fill in missing fields.

The questionnaire was developed and administered entirely in simplified Chinese, consistent with the native language of the target population—see Supplemental File for Chinese (in Multimedia Appendix 1) and translated English version (in Multimedia Appendix 2) of the questionnaire.

## Recruitment

The survey was carried out between December 16 and 30, 2022, among residents aged 18-69 years in Zhejiang province, China. Initially, an online open survey questionnaire was dispatched to the 11 municipal Centers for Disease Control and Prevention. These centers then actively promoted and conducted the online investigation by sharing survey links via local official WeChat public accounts, neighborhood WeChat groups, and other regional lifestyle apps. Visitors to the survey link could choose to exit at any time, and no access to other content was restricted to survey completion. The questionnaire was hosted on Wenjuan, a professional web-based platform specialized in questionnaires, tests, assessments, and voting. This platform is dedicated to offering users powerful and user-friendly services such as online questionnaire design, data collection, customized reports, and analysis of survey results.

## Measurements Control

### *Levels of Sleep Disturbance*

The Likert scale was used to assess the levels of sleep disturbance related to the COVID-19 epidemic reported by participants. In our research, we used a 5-point Likert scale which consists of 5 answer options: (1) not at all, (2) slight, (3) moderate, (4) significant, and (5) very significant.

### *Purchasing Behavior of COVID-19 Medicine*

Purchasing behavior of COVID-19 medicine was classified as follows: (1) participants who felt it unnecessary to purchase COVID-19 medicine, (2) participants who had tried, but were unable to access any COVID-19 medicine, and (3) participants who had successfully purchased COVID-19 medicine.

Participants were classified as having COVID-19 medication purchasing behavior during the early phase following China's zero-COVID policy adjustment if they had acquired the following drugs at that time: (1) cough medicine, (2) nasal congestion/snot relievers, (3) fever reducers, (4) pain relievers such as ibuprofen (Advil, Motrin IB, and others) or acetaminophen (Tylenol and others), and (5) other pharmaceuticals capable of alleviating COVID-19 infection symptoms.

### *Family Members Infected With COVID-19*

"Family members infected with COVID-19" was defined as "individuals who reside in the same household as the participant and had been diagnosed with COVID-19 (via nucleic acid or antigen testing) at any time prior to the survey."

## Statistical Analysis

Data were exported from the Wenjuan website to Excel (Microsoft), and were analyzed using the Statistical Analysis System (SAS), version SAS Viya Long-Term Support 2024.03 (SAS Institute Inc., Cary, NC, USA). Standard descriptive statistics were used for continuous and categorical variables to describe the characteristics of participants in this study. The $\chi^2$ test was used to explore the association between the purchasing behavior of COVID-19 medicines and the levels of sleep disturbance. Univariate logistic regression analysis was conducted to identify the sociodemographic factors associated with sleep disturbance. We used multivariate logistic regression to identify the association between sleep disturbance and purchasing behavior of COVID-19 medicines (categorical variable; reference group was "deemed unnecessary to purchase;" comparison groups was "successfully purchased" and "tried but failed to purchase"). The initial candidate variables included in the model were: age, gender, education level, occupation, marital status, and diagnosis of COVID-19 among family members. A backward stepwise selection approach was used, with a significance level of $P<.10$ for variable retention. Odds ratios (OR) with 95% CI were used to express measures of the associations. $P$ values less than .05 were considered to represent significance (two-sided).

## Ethical Considerations

The Zhejiang Provincial Center for Disease Control and Prevention Ethics Board approved the study protocol (2022-027-01). Conducted under strict anonymity, the survey safeguarded the privacy of all participants. No personally identifiable information was collected at any stage of the survey, the dataset was stored by specified staff, and all researchers involved signed a data confidentiality agreement. Informed consent was obtained from all participants prior to their inclusion in the study. By voluntarily clicking an electronic consent box, participants provided their informed consent, thereby indicating their willingness to participate in this research project. As compensation, participants could collect free health education materials at their local Center for Disease Control and Prevention.



XSL•FO
RenderX

## Results

### Respondents' Characteristics

A total of 40,130 residents took part in the survey, and after implementing quality control procedures, 1650 samples were excluded, with 38,480 samples ultimately included in the study. The sample consisted of 10,431 (27.11%) male participants and 28,049 (72.89%) female participants, with 1129 (2.93%) participants aged 20 years or below, 35,970 (93.48%) participants aged between 20 and 60, and 1384 (3.60%) participants over 60. Regarding occupations, 3049 (7.92%) were health workers, 7712 (20.04%) were government staff, 1522 (3.96%) were students , 1447 (3.76%) were farmers, and 24,750 (64.32%) were business/service industry workers and stay-at-home/unemployed individuals. In terms of education level, 22,180 (57.64%) participants had an education of junior college or undergraduate. Marital status showed that 31,053 (80.70%) people were married and 7427 (19.30%) people were unmarried, while 2873 (7.47%) people resided in households with confirmed COVID-19 cases. The detailed demographic characteristics of the participants are presented in Table 1.

**Table .** Sociodemographic characteristics of participants completing the online survey in the early stage after the lifting of zero-COVID policy in China (n=38,480).

| Characteristic | Count, n (%) |
|---|---|
| Age (y)[a] | |
| <20 | 1126 (2.93) |
| 20 - 60 | 35,970 (93.48) |
| >60 | 1384 (3.60) |
| Gender | |
| Male | 10,431 (27.11) |
| Female | 28,049 (72.89) |
| Occupation | |
| Health workers | 3049 (7.92) |
| Government staff | 7712 (20.04) |
| Students | 1522 (3.96) |
| Farmers | 1447 (3.76) |
| Business/service industry workers and stay-at-home/unemployed individuals | 24,750 (64.32) |
| Education level | |
| Primary school | 900 (2.34) |
| Junior high school | 7229 (18.79) |
| Senior high school | 7017 (18.24) |
| Junior college/undergraduate | 22,180 (57.64) |
| Postgraduate | 1154 (3) |
| Family members infected with COVID-19 | |
| Yes | 2873 (7.47) |
| No | 35,607 (92.53) |
| Marital status | |
| Married | 31,053 (80.70) |
| Others (single/divorced/widowed/separated, etc) | 7427 (19.30) |

[a]In the preliminary analysis, participants were divided into 10-year age groups (<20, 20 - 29, 30 - 39, 40 - 49, 50 - 59, 60 - 69 y). Given similar purchasing behavior in 20 - 29 to 50 - 59 years groups, these 4 groups were merged into 20 - 60 years for concise result presentation, with this data-driven rationale added herein.

### The Impact of China's Zero-COVID Policy Adjustment on the Sleep Status of the Public of Zhejiang, China

When the Likert scale was used to assess the degree of impact on participants' sleep, the findings revealed that out of 38,480 participants, 17,677 (45.94%) indicated that the adjustment of China's COVID-19 response policies had no effect on their sleep. In contrast, 20,803 (54.06%) reported experiencing varying levels of sleep disruption. Specifically, 10,964 (52.70%) described the impact as "slight," 3105 (14.93%) as "moderate," 3493 (16.79%) as "significant," and 3241 (15.58%) as "very significant."



XSL•FO
RenderX

### Association Between Sleep Status and Purchasing Behavior of COVID-19 Medicine Among the Public in the Early Stage After the Adjustment of China's Zero-COVID Policy

As shown in Table 2, among individuals who had tried, but were unable to access any COVID-19 medicine, 34.21% (7179/20,986) reported no impact on their sleep by the lifting of China's zero-COVID policy in the past week. In contrast, 54.81% (7538/13,752) of those who successfully acquired COVID-19 medications indicated no sleep disturbances. Notably, the proportion of individuals who deemed it unnecessary to buy COVID-19 medications and experienced no sleep disruption was as high as 79.10% (2960/3742).

**Table .** Impact of purchasing behavior of the COVID-19 medicine on the sleeping status among the participants of an online survey in the early stage after the lifting of zero-COVID policy in China.

| Purchasing of the COVID-19 medicine | How has your sleep been affected by the lifting of China's zero-COVID policy in the past week[a]? n (%) | | | | | Total |
|---|---|---|---|---|---|---|
| | Very significant | Significant | Moderate | Slight | Not at all | |
| Participants who felt it unnecessary to purchase COVID-19 medicine | 60 (1.60) | 62 (1.66) | 99 (2.65) | 561 (14.99) | 2960 (79.10) | 3742 |
| Participants who purchased COVID-19 medicine | 629 (4.57) | 931 (6.77) | 917 (6.67) | 3737 (27.17) | 7538 (54.81) | 13,752 |
| Participants who had tried, but were unable to access any COVID-19 medicine | 2552 (12.16) | 2500 (11.91) | 2089 (9.95) | 6666 (31.76) | 7179 (34.21) | 20,986 |

[a]$\chi^2$=3639.2; *P*<.001.

Among the 3742 respondents who considered purchasing COVID-19 medications unnecessary, the numbers of those experiencing slight, moderate, significant, and very significant impacts on sleep were 561 (14.99%), 99 (2.65%), 62 (1.66%), and 60 (1.60%), respectively. Among the 13,752 participants who had successfully obtained COVID-19 medications, the corresponding figures were 3737 (27.17%), 917 (6.67%), 931 (6.77%), and 629 (4.57%). For the 20,986 individuals who attempted to purchase but failed to obtain COVID-19 medications, the numbers were 6666 (31.76%), 2089 (9.95%), 2500 (11.91%), and 2552 (12.16%). A $\chi^2$ test for comparing multiple sample rates revealed that, overall, there were significant differences among the population rates (*P*<.001).

### Factors Associated With Sleep Disturbance Among the Public in the Early Stage After the Adjustment of China's Zero-COVID Policy

As shown in Table 3, there were statistically significant differences in sleep disturbance after the adjustment of China's zero-COVID policy among participants by age, education level, occupation, marital status, and diagnosis of COVID-19 among family members .



XSL•FO
RenderX

**Table .** Sociodemographic factors associated with sleep disturbance among the public in the early stage after the lifting of China's zero-COVID policy.

| Variables | Sleep status affected by the lifting of China's zero-COVID policy, n (%) | | OR[a] (95%CI) |
|---|---|---|---|
| | Yes | No | |
| Sex | | | |
| Male | 5716 (54.80) | 4715 (45.20) | 1.04 (0.99-1.09) |
| Female | 15,087 (53.79) | 12,962 (46.21) | Reference |
| Age (years) | | | |
| <20 | 409 (36.32) | 717 (63.68) | Reference |
| 20 - 60 | 19,714 (54.81) | 16,256 (45.19) | 2.13 (1.88-2.41)[b] |
| >60 | 680 (49.13) | 704 (50.87) | 1.69 (1.44-1.99)[b] |
| Education level | | | |
| Primary school | 517 (57.44) | 383 (42.56) | 1.36 (1.14-1.61)[b] |
| Junior high school | 3928 (54.34) | 3301 (45.66) | 1.19 (1.05-1.35)[c] |
| Senior high school | 3808 (54.27) | 3209 (45.73) | 1.18 (1.04-1.35)[c] |
| Junior college/undergraduate | 11,974 (53.99) | 10,206 (46.01) | 1.17 (1.03-1.31)[c] |
| Postgraduate | 576 (49.91) | 578 (50.09) | Reference |
| Occupation | | | |
| Health workers | 1242 (40.73) | 1807 (59.27) | Reference |
| Government staff | 4161 (53.095) | 3551 (46.05) | 1.71 (1.57-1.86)[b] |
| Students | 620 (40.74) | 902 (59.26) | 1.00 (0.88-1.13) |
| Business or service industry/stay-at-home/unemployed, etc | 13,925 (56.26) | 10,825 (43.74) | 1.87 (1.73-2.02)[b] |
| Farmers | 855 (59.09) | 592 (40.91) | 2.10 (1.85-2.39)[b] |
| Marital status | | | |
| Married | 17,143 (55.21) | 13,910 (44.79) | 1.27 (1.21-1.33)[b] |
| Others (divorced/widowed/separated, etc) | 3660 (49.28) | 3767 (50.72) | Reference |
| Has anyone in your family been diagnosed with COVID-19? | | | |
| Yes | 2023 (70.41) | 850 (29.59) | 2.13 (1.96-2.32)[b] |
| No | 18,780 (52.74) | 16,827 (47.26) | Reference |

[a]OR: odds ratio.

[b]$P<.001$.

[c]$P<.01$.

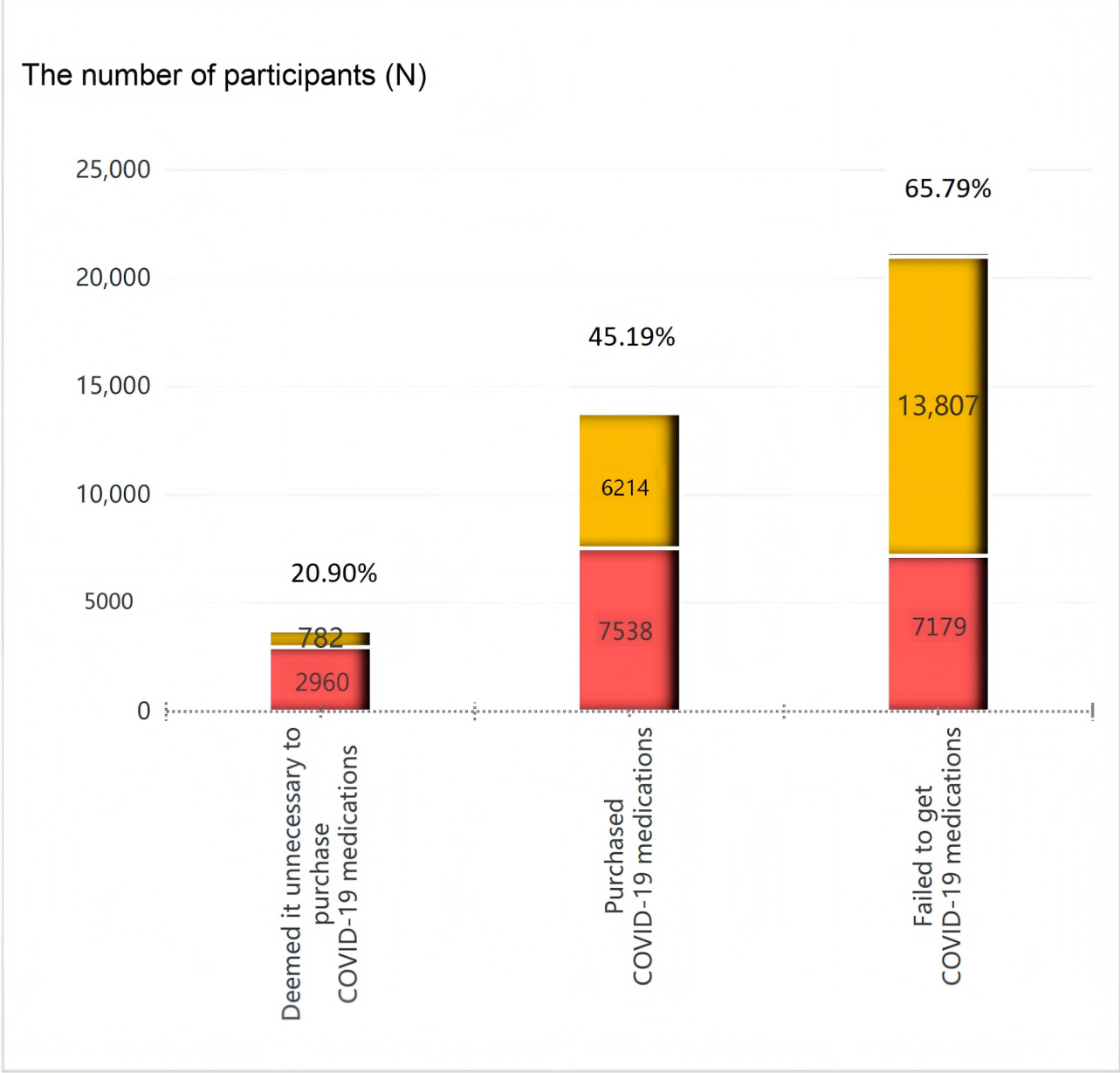
Age emerged as a notable predictor: participants aged 20 - 60 years had the highest sleep disturbance rate (19,714/35,970, 54.80%), followed by those over 60 years (680/1384, 49.13%), while 409 out of 1126 participants under 20 years reported sleep disturbance (409/1126, 36.32%). Education level showed a clear inverse relationship with sleep disturbance: the rate was highest among primary school graduates (517/900, 57.44%) and gradually decreased with higher education, reaching 49.91% (576/1154) for master's degree holders. Occupation also played a role. Farmers had the highest rate (855/1447, 59.09%), followed by business/service industry workers and stay-at-home/unemployed individuals (13,925/24,750, 56.26%), government staff (4161/7712, 53.95%), while 1242 out of 3049 health workers (1242/3049, 40.74%) and 620 out of 1522 students (620/1522, 40.74%) had lower rates. Marital status and family COVID-19 diagnosis further contributed to sleep disturbance: married participants had a higher (17,143/31,053, 55.21%) sleep disturbance rate than unmarried/separated individuals (OR 1.27, 95% CI 1.21 - 1.33). Most notably, participants with family members diagnosed with COVID-19 had the highest sleep disturbance rate (2023/2873, 70.41%) across all groups, with odds of disturbance 2.13 times higher than those without infected family members (OR 2.13, 95% CI 1.96 - 2.32) (detailed values in Table 3).



XSL•FO
RenderX

As shown in Figure 1, among individuals who had tried, but were unable to access any COVID-19 medicine, 65.79% (13,807/20,986) reported experiencing sleep disturbance. In contrast, 45.19% (6214/13,752) of those who successfully acquired COVID-19 medications indicated sleep disturbances. Notably, the proportion of individuals who deemed it unnecessary to buy COVID-19 medications and experienced sleep disruption was a mere 20.9% (782/3742).

**Figure 1.** Stacked bar chart showing individuals affected by sleep disturbance stratified by COVID-19 medication access status among the public in the early stage after the lifting of China's zero-COVID policy. The yellow color represents those who reported experiencing sleep disturbance, and the red color represents those unaffected.




By using multivariate logistic regression analysis and after adjusting for variables including gender, age, education level, occupation, marital status, and presence of family members diagnosed with COVID-19, it was discovered that there was a statistically significant correlation between purchasing behavior of COVID-19 medicine and the sleep status of the participants. Table 4 shows that, compared to those who deemed it unnecessary to buy COVID-19 medications, those who successfully acquired COVID-19 medications were 3.11 times more likely to encounter sleep disturbance (adjusted odds ratio [aOR] 3.11, 95% CI 2.85 - 3.39). Moreover, those who had tried, but were unable to access any COVID-19 medicine were 7.11 times (aOR 7.11, 95% CI 6.53 - 7.74) more likely to experience sleep disturbance.



XSL•FO
RenderX

**Table .** Multivariate logistic analysis on the association between the sleep status and the purchasing of COVID-19 medicines among the public in the early stage after the lifting of zero-COVID policy in China.

| Purchasing of COVID-19 medicine | Participants, n | Sleep status affected by COVID-19 epidemic, n (%) | aOR[a] (95% CI) | *P* value |
|---|---|---|---|---|
| Participants who felt it unnecessary to purchase COVID-19 medicine | 3742 | 782 (20.90) | Reference | —[b] |
| Participants who purchased COVID-19 medicine | 13,752 | 6214 (45.19) | 3.11 (2.85-3.39) | <.001 |
| Participants who had tried, but were unable to access any COVID-19 medicine | 20,986 | 13,807 (65.79) | 7.11 (6.53-7.74) | <.001 |

[a]aOR: adjusted odds ratio.

[b]Not applicable.

## Discussion

### Principal Findings

This study explores the sleep status of the public in Zhejiang province, China, following the adjustment of the zero-COVID policy and explores its relationship with the purchasing behavior towards COVID-19 medicine. The results offer valuable insights into the post-policy-adjustment situation. Given the dearth of similar studies, this research will serve as a foundation for future studies that highlight such crucial issues in the context of future public health emergencies.

After the zero-COVID policy adjustment, news and rumors about infection severity and fatalities—spread via social media, television, and daily interactions—fueled public anxiety and fear. Consequently, people act on their primal instincts and are prone to engage in panic buying to ease this psychological reaction [35]. The results of our study revealed that 35.74% (13,752/38,480) of the participants had successfully purchased COVID-19 medicines, 54.54% (20,986/38,480) had attempted to make a purchase but were unable to obtain any of the drugs, and 9.72% (3742/38,480) felt it was unnecessary to stockpile COVID-19 drugs. Psychologically, those who tried to purchase COVID-19 medicines may be driven by a fear of drug shortages, price increases, the surge in infections, and the availability of medicines and home delivery services during the pandemic [34,36]. Therefore, whether they successfully purchase medications directly affects their psychological state. Those who manage to buy medications will experience a certain degree of anxiety relief, while those who fail to do so may become more anxious [37-40].

Our study revealed that 54.06% (20,803/38,480) of the participants experienced sleep disturbances following the adjustment of the zero-COVID policy. It can be explained as a psychological burden during stressful circumstances. Pandemic-related sleep disturbance was also confirmed by many other studies worldwide. A study carried out in Beijing, China, found that, during the second wave of COVID-19, the overall prevalence of sleep disturbance was 50.8% [41]. Madeleine reported a 33.5% sleep disturbances during the COVID-19 shutdown phase in Germany [42]. It is also documented that there is a negative impact on the quality of sleep among the Spanish population during the lockdown period due to COVID-19 [43]. Studies found that COVID-19 had both negative and positive impacts on adolescent sleep. A study in Canada found that the shutdown of schools due to the COVID-19 pandemic might lead to a 2-hour shift in the sleep of typically developing adolescents, longer sleep duration, improved sleep quality, and less daytime sleepiness compared to those experienced under the regular school-time schedule [44]. Parents of 8th grade students from local schools across Ohio, Kentucky, and Virginia (USA) reported adolescents had more difficulties initiating and maintaining sleep during COVID-19 than before COVID-19, with clinically elevated rates increasing from 24% to 36%. Both bedtimes and wake times shifted later during COVID-19, and adolescents reported more delayed sleep/wake behaviors. Adolescents also reported less daytime sleepiness and longer school night sleep duration during COVID-19 [45].

Our study found that the degree of sleep disturbance varied, with 52.70% (10,964/20,803) reporting “slight,” 14.93% (3105/20,803) “moderate,” 16.79% (3493/20,803) “significant,” and 15.58% (3241/20,803) “very significant.” It suggests a complex and diverse range of effects on individuals, which may be related to various factors. Our study found that sociodemographic factors played a crucial role in sleep disturbances. Older people were more likely to experience sleep disturbances compared to the younger. A systematic review found that studies involving older participants had a higher pooled prevalence of insomnia symptoms during the COVID-19 pandemic [22]. Older individuals may be more sensitive to policy change impacts, due to preexisting health conditions, limited medical access, and greater reliance on stable daily routines—all of which increase their fear of infection [46]. Even though women were found to be more prone to sleep disturbance than men [47-49], the perceived impact of COVID-19 on sleep was not different between sexes in our study, which was also confirmed by research from the United States by Jeremy A Bigalke [50]. Literature reviews also revealed that sex had no bearing on the estimated prevalence of sleep disturbance [21]. Education level was inversely related to sleep disturbances; as the education level increased, the likelihood of experiencing sleep disturbances decreased. This could be because individuals with higher education levels are generally better at accessing and understanding information and medical resources, and thus better able to cope with the stress caused by the pandemic.



XSL•FO
RenderX

Studies found that people with higher socioeconomic status are more likely to attain medical resources during the COVID-19 pandemic [40,51]. While health care workers were documented as the most sleep-vulnerable group early in the COVID-19 pandemic [52,53], our study found the opposite in post-policy-adjustment in China. Different occupations face varying degrees of stress during the pandemic. For example, government staff may be burdened with epidemic-prevention work responsibilities, while business/service industry workers may worry about economic losses, and farmers may be concerned about the impact of the pandemic on agricultural production. However, at the very beginning after the adjustment of zero-COVID policy in China, the anxiety of getting infected without drugs in hand might relate to poor sleep quality. Our previous research found that 56.54% of health workers had successfully purchased COVID-19 medicines, which was the highest among all the occupation groups and might explain the less experience of sleep disturbance in this study [40]. Marital status and family members' COVID-19 diagnosis also had an impact. Married individuals had a higher probability of sleep disturbances, perhaps due to increased family responsibilities during the pandemic. Participants with family members diagnosed with COVID-19 were more prone to sleep disturbances, with a proportion as high as 70.41% (2023/2873), indicating that the direct impact of the disease on the family can cause significant psychological pressure, which was also confirmed by a study carried out in Bangladesh. Participants having relatives or acquaintances infected with COVID-19 were more likely to experience poor sleep quality (68.8% vs 49.0%) [47].

In addition to sociodemographic factors, our study found a strong association between the purchasing behavior of COVID-19 medicine and sleep status. Those who deemed purchasing unnecessary had the lowest proportion of sleep disruptions, while those who tried but failed to obtain medications had the highest. After adjusting for sociodemographic factors, it demonstrated that compared to those who thought purchasing was unnecessary, those who acquired medications were 3.11 times more likely, and those who tried but couldn't get medications were 7.11 times more likely to experience sleep disturbance. This difference highlights the role of medication availability and the psychological stress related to it [35]. For those who manage to buy the medicine, it may bring a sense of security and relieve some anxiety, as they feel prepared for potential infections. However, their higher sleep disturbance rate—compared to those who saw no need to purchase COVID-19 medicines—suggests lingering concerns about the disease and future drug shortages still impact their mental state. For those who fail to purchase it, the situation can exacerbate their anxiety. The fear of not being able to access necessary medications during illness and the uncertainty about their health can heighten stress levels, making them more likely to experience sleep disturbance. Our previous research revealed an inverse correlation between anxiety levels and the behavior of purchasing COVID-19 medications which can further explain the association between purchasing behavior of COVID-19 medicine and sleep status. Compared to those with severe anxiety, those with moderate anxiety were 1.76 times more likely to have purchased COVID-19 medicine (aOR 1.76, 95% CI 1.64 - 1.89); those with mild anxiety were 2.11 times (aOR 2.11, 95% CI 1.98 - 2.24) more likely to have purchased COVID-19 medicine; those with no anxiety were 2.48 times (aOR 2.48, 95% CI 2.31 - 2.67) more likely to have purchased COVID-19 medicine [40]. Individuals who were unable to obtain COVID-19 medications may have experienced heightened anxiety, which in turn could have contributed to increased instances of sleep disturbance. The very strong aOR of 7.11 for those unable to access medications and their likelihood of reporting sleep disturbances is notable. However, this causal inference requires further validation to account for unmeasured confounders (eg, baseline anxiety levels, preexisting sleep disorders).

## Limitations

This study has certain limitations. Firstly, to quickly explore the association between sleep disturbance and COVID-19 medicine purchasing behavior, the questionnaire was kept concise, omitting potential confounders (eg, COVID-19 vaccination status, comorbidities). "Participants who had tried, but were unable to access any COVID-19 medicine" was not clearly defined, which may lead to misclassification of participants' purchasing behavior. Sleep disturbance was measured via a single Likert-scale item—practical for large surveys but lacking details on sleep frequency/duration/severity, limiting study depth. These should be explored in future studies. Secondly, the data collection relied on self-reported information, which is subject to recall bias and may distort the study results. Thirdly, among the participants in the online survey, the participation of women and those with a higher education level is significantly higher than that of men and those with a low education level. Thus, applying the survey results directly to the general population could lead to inaccuracies.

## Conclusions

The findings reflect 54.1% of participants experiencing varying levels of sleep disturbance after the zero-COVID policy adjustment in Zhejiang, China. The older people, those with lower education levels, those with family members diagnosed with COVID-19, and those who had tried, but were unable to access any COVID-19 medicine were mostly affected. Public health agencies should design tailored communication campaigns for different demographic groups, especially for those high-risk groups, to minimize COVID-19-related fear and anxiety. Additionally, optimizing the supply and distribution of COVID-19 medications might help alleviate the stress of individuals who are unable to access them. A robust, transparent pharmaceutical supply chain monitoring system and dedicated channels (hotlines/online platforms) for reporting shortages would enable timely supply adjustments, ensuring high-risk groups access necessary drugs. Beyond China, our findings hold global relevance to observations in other countries where postpandemic policy shifts linked to anxiety and sleep disruptions [42,43].



XSL•FO
RenderX

## Acknowledgments

The authors are grateful to the Wenjuan platform that helped carry out the online survey and staff from the municipal Center for Disease Control and Prevention involved in this study.

## Data Availability

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to ethical restrictions.

## Authors' Contributions

Conceptualization: YH, X Zhao, SW
Data curation: SW, SX
Formal analysis: YH, SW, X Zhao
Investigation: YH, LW, QL, QW, SX
Supervision: SW, X Zhang
Writing—original draft: YH, X Zhao, SW
Writing—review and editing: YH, X Zhao, LW, QL, SW, QW, SX, X Zhang
All authors have read and agreed to the published version of the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire: Chinese version.
[DOC File, 31 KB - i-jmr_v15i1e79903_app1.doc ]

Multimedia Appendix 2
Questionnaire: translated English version.
[DOC File, 41 KB - i-jmr_v15i1e79903_app2.doc ]

## References

1. Hui DS, I Azhar E, Madani TA, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - the latest 2019 novel coronavirus outbreak in Wuhan, China. Int J Infect Dis 2020 Feb;91:264-266. [doi: 10.1016/j.ijid.2020.01.009] [Medline: 31953166]
2. Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan, China: the mystery and the miracle. J Med Virol 2020 Apr;92(4):401-402. [doi: 10.1002/jmv.25678] [Medline: 31950516]
3. Spina S, Marrazzo F, Migliari M, Stucchi R, Sforza A, Fumagalli R. The response of Milan's emergency medical system to the COVID-19 outbreak in Italy. Lancet 2020 Mar 14;395(10227):e49-e50. [doi: 10.1016/S0140-6736(20)30493-1] [Medline: 32119824]
4. Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Feb 10;41(2):145-151. [doi: 10.3760/cma.j.issn.0254-6450.2020.02.003] [Medline: 32064853]
5. Lim J, Jeon S, Shin HY, et al. Case of the index patient who caused tertiary transmission of coronavirus disease 2019 in Korea: the Application of Lopinavir/Ritonavir for the Treatment of COVID-19 pneumonia monitored by quantitative RT-PCR. J Korean Med Sci 2020 Feb 17;35(6):e79. [doi: 10.3346/jkms.2020.35.e79] [Medline: 32056407]
6. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review. JAMA 2020 Aug 25;324(8):782-793. [doi: 10.1001/jama.2020.12839] [Medline: 32648899]
7. Jernigan DB, CDC COVID-19 Response Team. Update: public health response to the coronavirus disease 2019 outbreak - United States, February 24, 2020. MMWR Morb Mortal Wkly Rep 2020 Feb 28;69(8):216-219. [doi: 10.15585/mmwr.mm6908e1] [Medline: 32106216]
8. Eurosurveillance editorial team. Note from the editors: World Health Organization declares novel coronavirus (2019-nCoV) sixth public health emergency of international concern. Euro Surveill 2020 Feb;25(5):200131e. [doi: 10.2807/1560-7917.ES.2020.25.5.200131e] [Medline: 32019636]
9. Tang S, Wang X, Tang B, et al. Threshold conditions for curbing COVID-19 with a dynamic zero-case policy derived from 101 outbreaks in China. BMC Public Health 2023 Jun 6;23(1):1084. [doi: 10.1186/s12889-023-16009-8] [Medline: 37280554]



XSL•FO
RenderX

10. Li H, Zhang H. Cost-effectiveness analysis of COVID-19 screening strategy under China's dynamic zero-case policy. Front Public Health 2023;11:1099116. [doi: 10.3389/fpubh.2023.1099116] [Medline: 37228729]
11. China's response to COVID-19: a chance for collaboration. Lancet 2021 Apr 10;397(10282):1325. [doi: 10.1016/S0140-6736(21)00823-0] [Medline: 33838746]
12. Huang Y, Wu Q, Wang P, et al. Measures undertaken in China to avoid COVID-19 infection: internet-based, cross-sectional survey study. J Med Internet Res 2020 May 12;22(5):e18718. [doi: 10.2196/18718] [Medline: 32396516]
13. Wei Y, Ye Z, Cui M, Wei X. COVID-19 prevention and control in China: grid governance. J Public Health (Oxf) 2021 Apr 12;43(1):76-81. [doi: 10.1093/pubmed/fdaa175]
14. Wilson O, Flahault A. China's U-turn in its COVID-19 policy. Anaesth Crit Care Pain Med 2023 Apr;42(2):101197. [doi: 10.1016/j.accpm.2023.101197] [Medline: 36646356]
15. Notice on further optimizing and implementing prevention and control measures for COVID-19 [Web page in Chinese]. National Health Commission of the People's Republic of China. 2022. URL: https://www.nhc.gov.cn/xcs/zhengcwj/202211/c95de5f0be9b4e49b99517385aa662a0.shtml [accessed 2025-12-12]
16. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey. Psychiatry Res 2020 Jun;288:112954. [doi: 10.1016/j.psychres.2020.112954] [Medline: 32325383]
17. Partinen M, Bjorvatn B, Holzinger B, et al. Sleep and circadian problems during the coronavirus disease 2019 (COVID-19) pandemic: the International COVID-19 Sleep Study (ICOSS). J Sleep Res 2021 Feb;30(1):e13206. [doi: 10.1111/jsr.13206] [Medline: 33179820]
18. Dongol E, Shaker K, Abbas A, et al. Sleep quality, stress level and COVID-19 in university students; the forgotten dimension. Sleep Sci 2022;15(Spec 2):347-354. [doi: 10.5935/1984-0063.20210011] [Medline: 35371400]
19. Jahrami H, BaHammam AS, Bragazzi NL, Saif Z, Faris M, Vitiello MV. Sleep problems during the COVID-19 pandemic by population: a systematic review and meta-analysis. J Clin Sleep Med 2021 Feb 1;17(2):299-313. [doi: 10.5664/jcsm.8930] [Medline: 33108269]
20. Magklara K, Kyriakopoulos M. The impact of the COVID-19 pandemic on children and young people. Psychiatriki 2023 Dec 29;34(4):265-268. [doi: 10.22365/jpsych.2023.024] [Medline: 37982248]
21. Jahrami HA, Alhaj OA, Humood AM, et al. Sleep disturbances during the COVID-19 pandemic: a systematic review, meta-analysis, and meta-regression. Sleep Med Rev 2022 Apr;62:101591. [doi: 10.1016/j.smrv.2022.101591] [Medline: 35131664]
22. Li Y, Chen B, Hong Z, et al. Insomnia symptoms during the early and late stages of the COVID-19 pandemic in China: a systematic review and meta-analysis. Sleep Med 2022 Mar;91:262-272. [doi: 10.1016/j.sleep.2021.09.014] [Medline: 34732293]
23. Kumar N, Gupta R. Disrupted sleep during a pandemic. Sleep Med Clin 2022 Mar;17(1):41-52. [doi: 10.1016/j.jsmc.2021.10.006] [Medline: 35216760]
24. Krueger JM, Frank MG, Wisor JP, Roy S. Sleep function: toward elucidating an enigma. Sleep Med Rev 2016 Aug;28:46-54. [doi: 10.1016/j.smrv.2015.08.005] [Medline: 26447948]
25. Vyazovskiy VV. Sleep, recovery, and metaregulation: explaining the benefits of sleep. Nat Sci Sleep 2015;7:171-184. [doi: 10.2147/NSS.S54036] [Medline: 26719733]
26. Richter K, Kellner S, Hillemacher T, Golubnitschaja O. Sleep quality and COVID-19 outcomes: the evidence-based lessons in the framework of predictive, preventive and personalised (3P) medicine. EPMA J 2021 Jun;12(2):221-241. [doi: 10.1007/s13167-021-00245-2] [Medline: 34122671]
27. Liew SC, Aung T. Sleep deprivation and its association with diseases- a review. Sleep Med 2021 Jan;77:192-204. [doi: 10.1016/j.sleep.2020.07.048] [Medline: 32951993]
28. Cox RC, Olatunji BO. Sleep in a pandemic: Implications of COVID-19 for sleep through the lens of the 3P model of insomnia. Am Psychol 2021 Oct;76(7):1159-1171. [doi: 10.1037/amp0000850] [Medline: 34138586]
29. Papagiouvanni I, Kotoulas SC, Vettas C, Sourla E, Pataka A. Sleep during the COVID-19 pandemic. Curr Psychiatry Rep 2022 Nov;24(11):635-643. [doi: 10.1007/s11920-022-01371-y] [Medline: 36192579]
30. Ara T, Rahman MM, Hossain MA, Ahmed A. Identifying the associated risk factors of sleep disturbance during the COVID-19 lockdown in Bangladesh: a web-based survey. Front Psychiatry 2020;11:580268. [doi: 10.3389/fpsyt.2020.580268] [Medline: 33093839]
31. Das S, Khan R, Banerjee S, Ray S, Ray S. Alterations in circadian rhythms, sleep, and physical activity in COVID-19: mechanisms, interventions, and lessons for the future. Mol Neurobiol 2024 Dec;61(12):10115-10137. [doi: 10.1007/s12035-024-04178-5] [Medline: 38702566]
32. Lissak G. Adverse physiological and psychological effects of screen time on children and adolescents: literature review and case study. Environ Res 2018 Jul;164:149-157. [doi: 10.1016/j.envres.2018.01.015] [Medline: 29499467]
33. Al Zoubi S, Gharaibeh L, Jaber HM, Al-Zoubi Z. Household drug stockpiling and panic buying of drugs during the COVID-19 pandemic: a study from Jordan. Front Pharmacol 2021;12:813405. [doi: 10.3389/fphar.2021.813405] [Medline: 35002744]




XSL•FO
RenderX

34. Miah MS, Mamun MR, Saif-Ur-Rahman KM, Rabby AA, Zakaria AFM. A qualitative exploration of purchasing, stockpiling, and use of drugs during the COVID-19 pandemic in an urban city of Bangladesh. Public Health Pract (Oxf) 2024 Jun;7:100477. [doi: 10.1016/j.puhip.2024.100477] [Medline: 38379753]
35. Arafat SMY, Kar SK, Marthoenis M, Sharma P, Hoque Apu E, Kabir R. Psychological underpinning of panic buying during pandemic (COVID-19). Psychiatry Res 2020 Jul;289:113061. [doi: 10.1016/j.psychres.2020.113061] [Medline: 32413711]
36. Jifar WW, Geneti GB, Dinssa SD. The impact of COVID-19 on pharmaceutical shortages and supply disruptions for non-communicable diseases among public hospitals of South West, Oromia, Ethiopia. J Multidiscip Healthc 2022;15:1933-1943. [doi: 10.2147/JMDH.S377319] [Medline: 36072279]
37. Sari E, Kağan G, Karakuş B, Özdemir Ö. Dataset on social and psychological effects of COVID-19 pandemic in Turkey. Sci Data 2022 Jul 23;9(1):441. [doi: 10.1038/s41597-022-01563-4] [Medline: 35871225]
38. Taylor S. Understanding and managing pandemic-related panic buying. J Anxiety Disord 2021 Mar;78:102364. [doi: 10.1016/j.janxdis.2021.102364] [Medline: 33517219]
39. Goodwin R, Wiwattanapantuwong J, Tuicomepee A, Suttiwan P, Watakakosol R, Ben-Ezra M. Anxiety, perceived control and pandemic behaviour in Thailand during COVID-19: results from a national survey. J Psychiatr Res 2021 Mar;135:212-217. [doi: 10.1016/j.jpsychires.2021.01.025] [Medline: 33497875]
40. Huang Y, Xu S, Zhao X, et al. Drug stockpiling behavior and its impact on anxiety among the general public in the early stage after the lifting of China's zero-COVID policy: results from a web-based survey. Front Pharmacol 2025;16:1524068. [doi: 10.3389/fphar.2025.1524068] [Medline: 40421214]
41. Liu Y, Wang X, Sun P, et al. Sleep disturbance and anxiety symptom among public during the second wave of COVID-19 in Beijing: a web-based cross-sectional survey. J Affect Disord 2022 Feb;298(Pt A):80-85. [doi: 10.1016/j.jad.2021.10.068] [Medline: 34728284]
42. Hetkamp M, Schweda A, Bäuerle A, et al. Sleep disturbances, fear, and generalized anxiety during the COVID-19 shut down phase in Germany: relation to infection rates, deaths, and German stock index DAX. Sleep Med 2020 Nov;75:350-353. [doi: 10.1016/j.sleep.2020.08.033] [Medline: 32950879]
43. Maestro-Gonzalez A, Sánchez-Zaballos M, Mosteiro-Díaz MP, Zuazua-Rico D. Quality of sleep among social media users during the lockdown period due to COVID-19 in Spain. Sleep Med 2021 Apr;80:210-215. [doi: 10.1016/j.sleep.2021.01.050] [Medline: 33607556]
44. Gruber R, Saha S, Somerville G, Boursier J, Wise MS. The impact of COVID-19 related school shutdown on sleep in adolescents: a natural experiment. Sleep Med 2020 Dec;76:33-35. [doi: 10.1016/j.sleep.2020.09.015] [Medline: 33070000]
45. Becker SP, Dvorsky MR, Breaux R, Cusick CN, Taylor KP, Langberg JM. Prospective examination of adolescent sleep patterns and behaviors before and during COVID-19. Sleep 2021 Aug 13;44(8):zsab054. [doi: 10.1093/sleep/zsab054] [Medline: 33631014]
46. Du M, Liu M, Wang Y, Qin C, Liu J. Global burden of sleep disturbances among older adults and the disparities by geographical regions and pandemic periods. SSM Popul Health 2024 Mar;25:101588. [doi: 10.1016/j.ssmph.2023.101588] [Medline: 38225953]
47. Islam MS, Rahman ME, Zubayer AA, et al. Investigating poor sleep quality and associated factors during the COVID-19 pandemic: a population-based survey in Bangladesh. Front Public Health 2021;9:724520. [doi: 10.3389/fpubh.2021.724520] [Medline: 34881215]
48. Mandelkorn U, Genzer S, Choshen-Hillel S, et al. Escalation of sleep disturbances amid the COVID-19 pandemic: a cross-sectional international study. J Clin Sleep Med 2021 Jan 1;17(1):45-53. [doi: 10.5664/jcsm.8800] [Medline: 32900428]
49. Casagrande M, Favieri F, Tambelli R, Forte G. The enemy who sealed the world: effects quarantine due to the COVID-19 on sleep quality, anxiety, and psychological distress in the Italian population. Sleep Med 2020 Nov;75:12-20. [doi: 10.1016/j.sleep.2020.05.011] [Medline: 32853913]
50. Bigalke JA, Greenlund IM, Carter JR. Sex differences in self-report anxiety and sleep quality during COVID-19 stay-at-home orders. Biol Sex Differ 2020 Oct 13;11(1):56. [doi: 10.1186/s13293-020-00333-4] [Medline: 33050941]
51. Elek P, Bíró A, Fadgyas-Freyler P. Income gradient of pharmaceutical panic buying at the outbreak of the COVID-19 pandemic. Health Econ 2021 Sep;30(9):2312-2320. [doi: 10.1002/hec.4378] [Medline: 34218496]
52. Wu K, Wei X. Analysis of psychological and sleep status and exercise rehabilitation of front-line clinical staff in the fight against COVID-19 in China. Med Sci Monit Basic Res 2020 May 11;26:e924085. [doi: 10.12659/MSMBR.924085] [Medline: 32389999]
53. Zhang C, Yang L, Liu S, et al. Survey of insomnia and related social psychological factors among medical staff involved in the 2019 novel coronavirus disease outbreak. Front Psychiatry 2020;11:306. [doi: 10.3389/fpsyt.2020.00306] [Medline: 32346373]


## Abbreviations

**aOR:** adjusted odds ratio
**OR:** odds ratio
**SAS:** Statistical Analysis System



XSL•FO
RenderX

*Edited by N Cahill; submitted 30.Jun.2025; peer-reviewed by MC Yang, OC Balogun; accepted 28.Oct.2025; published 06.Jan.2026.*

*Please cite as:*
*Huang Y, Zhao X, Wang L, Lv Q, Wu Q, Xu S, Zhang X, Wu S*
*Sleep Disturbance and Its Association With Purchasing Behavior of COVID-19 Medicine Among the Public After the Adjustment of Zero-COVID Policy in China: Results From a Web-Based Survey Study*
*Interact J Med Res 2026;15:e79903*
*URL: https://www.i-jmr.org/2026/1/e79903*
*doi:10.2196/79903*

© Yu Huang, Xiang Zhao, Lei Wang, Qiaohong Lv, Qingqing Wu, Shuiyang Xu, Xuehai Zhang, Shuxian Wu. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 6.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Clinical Effectiveness and Cost-Effectiveness of Blended Treatment for Major Depression Compared With Treatment as Usual Within Routine Care in Europe: A Noninferiority Randomized Controlled Trial

Heleen Riper[1,2], PhD; Annet Kleiboer[1], PhD; Adriaan Hoogendoorn[2], PhD; Judith Bosmans[3], PhD; Mieke Schulte[1,4], PhD; Kim Mathiasen[5], PhD; Jean Baptiste Hazo[6,7], PhD; Jerome Holtzmann[8,9], PhD; Karine Chevreul[6,7], PhD; David Daniel Ebert[10], PhD; Ingrid Titzler[11,12], PhD; Mathias Berking[11], PhD; Burkhardt Funk[13], PhD; Anneke van Schaik[2], PhD; Lise Kemmeren[14], PhD; Roman Cieslak[15,16], PhD; Ewelina Smoktunowicz[17], PhD; Anna Maj[17], PhD; Cristina Botella[18,19], PhD; Azucena Garcia-Palacios[18,19], PhD; Rosa Banos[19,20], PhD; Rocío Herrero[19,20], PhD; Gerhard Andersson[21,22,23], PhD; Naira Topooco[21,24], PhD; Kristofer Vernmark[21], PhD; Thomas Berger[25], PhD; Tobias Krieger[25], PhD; Arlinda Cerga-Pashoja[26,27], PhD; Asmae Doukani[26], PhD; Ricardo Araya[28], PhD; Eirini Karyotaki[1,29], PhD; Spyros Kolovos[3], PhD; Mark Hoogendoorn[30], PhD; Ward van Breda[30], PhD; Pepijn van de Ven[31], PhD; Artur Rocha[32], PhD; Gonçalo Gonçalves[32], MSc; Pim Cuijpers[1], PhD

[1]Department of Clinical, Neuro- and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Van der Boechorststraat 7, Amsterdam, The Netherlands

[10]Psychology and Digital Mental Health Care, TUM School of Medicine and Health, Technical University of Munich, Munich, Germany

[11]Department of Clinical Psychology and Psychotherapy, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

[12]Institute of Psychology, UMIT TIROL, Private University for Health Sciences and Technology, Hall in Tirol, Austria

[13]Institute of Information Systems, Leuphana University, Lüneburg, Germany

[14]Department of Research and Innovation, GGZ InGeest Specialized Mental Health Care, Amsterdam, The Netherlands

[15]Faculty of Psychology, SWPS University, Warsaw, Poland

[16]Lyda Hill Institute for Human Resilience, University of Colorado Colorado Springs, Colorado Springs, CO, United States

[17]StresLab Research Centre, Institute of Psychology, SWPS University, Warsaw, Poland

[18]Department of Basic Psychology, Clinical and Psychobiology, Universitat Jaume I, Castellón de la Plana, Spain

[19]CIBERobn Physiopathology of Obesity and Nutrition, Instituto de Salud Carlos III, Madrid, Spain

[2]Amsterdam UMC, Location Vrije Universiteit Amsterdam, Department of Psychiatry, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

[20]Department of Personality, Evaluation and Psychological Treatments, Universidad de Valencia, Valencia, Spain

[21]Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden

[22]Lusófona University, HEI-Lab: Digital Human-EnvironmentInteraction Labs, Lisboa, Portugal

[23]Department of Health, Education and Technology, Luleå University of Technology, Luleå, Sweden

[24]Department of Psychology, Uppsala University, Uppsala, Sweden

[25]Department of Clinical Psychology and Psychotherapy, University of Bern, Bern, Switzerland

[26]Department of Population Health, London School of Hygiene & Tropical Medicine, London, United Kingdom

[27]School of Allied Health and Life Sciences, Faculty of Sport, Technology and Health Sciences, St Mary's University, Twickenham, United Kingdom

[28]Department of Health Service and Population Research, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[29]Department of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands

[3]Department of Health Sciences, Faculty of Earth and Life Sciences, Amsterdam Public Health Research Institute, Vrije Universiteit, Amsterdam, The Netherlands

[30]Department of Computer Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[31]Health Research Institute, University of Limerick, Limerick, Ireland

[32]Human-Centered Computing and Information Science, Institute for Systems Engineeringand Computers Technology and Science (INESCTEC), Porto, Portugal

[4]Department of Psychology, Education and Child Studies, Clinical Psychology, Erasmus School of Social and Behavioral Sciences, Erasmus University Rotterdam, Rotterdam, The Netherlands

[5]Department of Psychology and Behavioral Sciences, Aarhus University, Aarhus, Denmark

[6]ECEVE, UMR 1123, Inserm, Université Paris Cité, Paris, France

[7]Health Economics Research Unit, Inserm, University of Paris, Paris, France

[8]Service de Psychiatrie de l'Adulte, Centre Expert Dépression Résistante FondaMental, CHU Grenoble Alpes, Hôpital Nord, Grenoble, France



XSL•FO
RenderX

[9]Fondation FondaMental (FondaMental Foundation), Créteil, France

**Corresponding Author:**
Heleen Riper, PhD
Department of Clinical, Neuro- and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Van der Boechorststraat 7, Amsterdam, The Netherlands

## Abstract

**Background:** Cognitive behavioral therapy (CBT) is an effective and widely used treatment for major depressive disorder (MDD). However, access is limited by long waiting lists and a shortage of trained therapists. Evidence-based guided and self-guided digital interventions can address these challenges, but their large-scale adoption in routine primary and specialized mental health care has been slow. Blended CBT (bCBT), combining face-to-face therapy with structured, guided digital treatment modules, may increase treatment capacity while maintaining the benefits of therapist support.

**Objective:** The European Comparative Effectiveness Research on Internet-based Depression Treatment study, conducted across 9 European countries, is the first large-scale comparative bCBT study for MDD. The hypothesis was that bCBT is clinically noninferior and cost-effective when compared with treatment as usual (TAU), which mainly consisted of face-to-face CBT.

**Methods:** A multisite randomized controlled trial was conducted with a noninferiority margin of *d*=0.20. Main inclusion criteria were age ≥18 years, a diagnosis of MDD based on the Mini International Neuropsychiatric Interview, and a baseline Patient Health Questionnaire-9 (PHQ-9) score of ≥5. The primary outcome was the PHQ-9, with secondary outcomes including MDD remission at 12 months, therapeutic alliance, and costs. Intention-to-treat analyses were performed, and linear mixed modeling was used to assess intervention effects. Cost-effectiveness analyses were conducted from a health care perspective.

**Results:** A total of 835 patients were included in the study. bCBT was shown to be noninferior to TAU on the primary PHQ-9 outcome during treatment (3 months after the baseline assessment, *d*=−0.26, 95% CI −0.43 to −0.09), at 6 months (posttreatment, *d*=−0.21, 95% CI −0.39 to−0.04), and at 12-month follow-up (*d*=−0.01, 95% CI −0.19 to 0.17). Furthermore, the bCBT group had a significantly lower likelihood of experiencing an MDD episode at 12 months (odds ratio 0.67, 95% CI 0.45 - 0.99). Subgroup analyses indicated that participants with and with no antidepressant use at baseline in both groups benefited equally from their treatment. Deterioration rates (Reliable Change Index) were below 5% in both groups. bCBT appeared acceptable for patients and therapists, with a strong working alliance in both groups. From a health care perspective, bCBT was not cost-effective at 12 months' assessment, as its total costs were not significantly lower than TAU. However, the probability that bCBT is cost-effective is high (0.95) at a willingness to pay (WTP) of €3800 per improvement in the PHQ-9 score, and 0.76 at a WTP of €10,000 per MDD case prevented (average 2017 exchange rate was €1=US $1.13).

**Conclusions:** bCBT offers an effective and safe digitally supported alternative to face-to-face TAU for patients with MDD and their therapists in routine clinical care. From a health care perspective, its cost-effectiveness depends on policymakers' WTP for the additional clinical benefits achieved.

**Trial Registration:** ClinicalTrials.gov NCT02796573; https://clinicaltrials.gov/study/NCT02796573 and ClinicalTrials.gov NCT02542891; https://clinicaltrials.gov/study/NCT02542891 and German Clinical Trials Register DRKS00006866; https://drks.de/search/en/trial/DRKS00006866 and Netherlands Trials Register NTR4962; https://www.onderzoekmetmensen.nl/en/trial/25452 and ClinicalTrials.gov NCT02389660; https://clinicaltrials.gov/study/NCT02389660 and ClinicalTrials.gov NCT02361684; https://clinicaltrials.gov/study/NCT02361684 and ClinicalTrials.gov NCT02449447; https://clinicaltrials.gov/study/NCT02449447 and ClinicalTrials.gov NCT02410616; https://clinicaltrials.gov/study/NCT02410616 and ISRCTN Registry ISRCTN12388725; https://www.isrctn.com/ISRCTN12388725

**International Registered Report Identifier (IRRID):** RR2-10.1186/s13063-016-1511-1

***(Interact J Med Res 2026;15:e72866)*** doi:10.2196/72866

**KEYWORDS**

major depression disorder; cognitive behavioral therapy; blended treatment; cost-effectiveness; noninferiority randomized controlled trial; digital mental health

## Introduction

Globally, major depressive disorder (MDD) is among the most prevalent mental disorders, affecting approximately 4% of the population each year, corresponding to an estimated 332 million people in 2021 [1]. MDD is a leading cause of disability worldwide and imposes a substantial societal burden through increased health care usage and major productivity losses [2,3]. MDD is frequently accompanied by psychiatric comorbidities, particularly anxiety disorders and alcohol use disorders [4].



XSL•FO
RenderX

Although antidepressants (ADs), psychotherapy, and their combination are all effective treatments for MDD in the short- and long-term [5], patients often prefer psychotherapy over ADs, partly because of concerns about adverse effects and a greater sense of autonomy and self-management [6].

Cognitive behavioral therapy (CBT) is one of the most extensively studied and widely implemented evidence-based psychotherapies for MDD in primary and specialized mental health care [5]. Compared with treatment as usual (TAU) or waiting-list control conditions, CBT demonstrates moderate to large treatment effects, whereas comparisons with other evidence-based psychotherapies generally show only small or nonsignificant differences [5]. Despite this strong evidence base, scaling up face-to-face CBT (ftfCBT) in routine care remains challenging. High treatment costs, long waiting lists, shortages of trained therapists, and limited reimbursement for CBT services—even in high-income countries—continue to restrict access to care [7]. As a result, a substantial global treatment gap persists for people with MDD [8].

Over the past 2 decades, internet-based cognitive behavioral therapy (iCBT) has been developed to address some of these challenges by offering an accessible, cost-effective, and scalable treatment format [9,10]. Numerous meta-analyses have demonstrated that self-guided and therapist-guided iCBT are effective in treating depression. Compared with nonactive controls and participants recruited from the general population, therapist-guided iCBT generally yields moderate to large effect sizes, whereas self-guided iCBT produces smaller, but still significant, effects [11]. Furthermore, a meta-analysis found no significant differences in depression outcomes between iCBT and ftfCBT among patients receiving routine mental health care. However, this meta-analysis by Hedman-Lagerlöf and colleagues [12] included only 5, mostly small-scale, randomized controlled trials (RCTs), highlighting the limited evidence available from routine care settings. Although these findings support the efficacy of iCBT, evidence from routine clinical practice remains limited. Together with the scarcity of implementation studies conducted under real-world conditions, this may partly explain the relatively slow uptake of iCBT for MDD in primary and specialized mental health care (with the exception of dedicated online treatment services).

In addition to the limited routine care evidence base, several barriers have been proposed to explain the slow implementation of iCBT. These include organizational factors (eg, limited managerial support), therapist-related concerns, patient-related barriers, and external factors such as data privacy regulations [13]. For example, therapists question whether findings from existing iCBT studies are generalizable to patients treated in routine clinical practice, who are often perceived to be more clinically complex and socially vulnerable than participants recruited from the general population because of greater symptom severity, higher rates of comorbidity, and greater socioeconomic disadvantage [14]. Therapists with little or no experience using iCBT often worry as well that delivering treatment remotely will hinder the development of a strong therapeutic alliance, despite evidence indicating that therapeutic alliance can be maintained [15,16]. Patients have identified barriers, also including the high level of motivation required to work independently and the reduced amount of face-to-face contact.

These implementation barriers have stimulated interest in blended treatment formats for MDD in routine care. Reviews have shown that blended CBT (bCBT) protocols vary with respect to their treatment composition (integrated or sequential), the role of the digital component (core or supplementary), and the technologies applied (internet, mobile apps, or both) [17,18]. By combining the strengths of face-to-face and digital treatment, bCBT may overcome several perceived limitations of stand-alone iCBT while retaining its advantages in terms of accessibility and efficiency. Recent studies have explored the acceptability and preferences of therapists and patients for bCBT in routine care. These studies indicate that many therapists and patients prefer bCBT over stand-alone iCBT if bCBT maintains the advantage of the clinical flexibility of face-to-face therapy and digitally supports patients' self-management skills [19,20]. Few studies have evaluated the effectiveness of bCBT yet.

Nakao et al (2018) evaluated the effectiveness of bCBT in a small RCT (n=40), comparing bCBT plus standard psychiatric care with standard psychiatric care alone (waitlist) in patients with MDD who had not responded to AD medication. The study reported a large between-group effect in favor of bCBT ($d$=1.1) [21]. A smaller number of studies have yet clinically evaluated bCBT directly compared with ftfCBT for MDD. Kalde et al [22] (n=82) and Kooistra et al [23] (n=102), for example, reported no significant differences in depressive symptoms between treatment conditions at posttreatment. Although these findings are encouraging, neither study was powered to establish noninferiority. In addition, several bCBT trials have been described in published study protocols, with results awaited [24-26]. Consequently, robust evidence from large, pragmatic, multinational trials evaluating the clinical effectiveness and cost-effectiveness of bCBT in routine clinical practice remains limited.

To address this evidence gap, we conducted the European Comparative Effectiveness Research on Internet-based Depression Treatment (E-COMPARED) study across 9 European countries, evaluating an integrated form of bCBT in which face-to-face sessions were combined with structured internet- and smartphone-based modules within a single treatment protocol [27]. Within this protocol, face-to-face sessions were used for more in-depth therapeutic interactions, assessment of clinical needs, and shared decision-making, whereas the digital modules supported psychoeducation, homework assignments, symptom monitoring, and self-management.

We hypothesized that bCBT would be clinically noninferior to TAU for patients with MDD. In addition, we hypothesized that bCBT would reduce health care costs by requiring less therapist time while maintaining comparable clinical outcomes, thereby improving cost-effectiveness from a health care perspective. Demonstrating that bCBT is clinically noninferior and cost-effective would support its wider implementation, thereby improving access to evidence-based CBT for patients with MDD.



XSL•FO
RenderX

## Methods

### Study Design and Participants

We conducted a pragmatic 2-arm, open-label, multisite, randomized controlled clinical trial. We applied noninferiority design criteria, including an active comparator (TAU) and a preset noninferiority margin [28]. The E-COMPARED study was conducted across 8 European countries: France, Germany, the Netherlands, Poland, Spain, Sweden, Switzerland, and the United Kingdom, ensuring geographical coverage across Europe. Denmark was subsequently added due to the country's emergent interest in bCBT. We conducted the RCTs in settings where both bCBT and TAU for MDD treatment could be provided. In Denmark, France, the Netherlands, and Switzerland, the study was conducted within specialized mental health services, while in Germany, Poland, Spain, Sweden, and the United Kingdom, it was conducted within primary care. The results were reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) guidelines for noninferiority trials [29]. We refer to the protocol for a detailed description of the study [27].

### Inclusion and Exclusion Criteria

The inclusion criteria were age ≥18 years, meeting *DSM-IV* (*Diagnostic and Statistical Manual of Mental Disorders* [Fourth Edition]) diagnostic criteria for MDD via the Mini International Neuropsychiatric Interview (MINI; excluding sections M [anorexia nervosa], N [bulimia nervosa], and P [antisocial personality disorder]) [30], and depressive symptoms (Patient Health Questionnaire-9 [PHQ-9] score ≥5) [31]. Patients were permitted to receive pharmacotherapy in addition to bCBT at baseline or during treatment, if deemed necessary by a health care professional, in line with routine clinical practice. At the 12-month follow-up, phone-based MINI interviews were conducted by independent, trained interviewers who were masked to the assigned group. The exclusion criteria included currently being at high risk of suicide (MINI section C), psychiatric comorbidity (substance dependence, bipolar disorder, psychosis, and obsessive-compulsive disorder), currently undergoing psychological treatment for depression, language comprehension barriers, lack of a stable internet connection or access to a computer, and either the absence of a compatible smartphone or an unwillingness to loan one for bCBT.

### Procedures

Recruitment took place between February 2015 and December 2016; data collection closed on December 31, 2017. Patients were approached by mental health professionals in health care settings, who then referred them to the research team for eligibility assessment. In Germany and Sweden, recruitment leaflets were placed in general practitioner waiting rooms. Sweden also used social media platforms to recruit patients who had contacted primary care services about depressive complaints. Multimedia Appendix 1 provides an overview of the recruitment procedures for each country, along with any additional procedures carried out.

### Ethical Considerations

Ethical approval and trial registrations were obtained at the national level. Participants provided written informed consent, authorizing the E-COMPARED partners to share and analyze the anonymized dataset and to report the results in publications. Participants did not receive compensation for study participation. Ethical approval for the trials has been obtained locally in each country: Ethics Committee of the Region of Southern Denmark (S-20150150); France: Comité de protection des personnes, Île de France V (15033-n 2015-A00565-44); Germany: Ethik Kommission DGPsychologie, Universität Trier (MB 102014); The Netherlands: METC VUMC (2015.078); Poland: Komisja ds. Etyki Badań Naukowych (10/2014); Spain: Comisión Deontológica/Comité Ético de Investigación en Humanos (H1414775276823); Sweden: Regionala etikprövningsnämnden (2014/428-31); Switzerland: Kantonale Ethikkommission Bern (001/2015); and United Kingdom: NRES Committee London-Camden and King's Cross (15/LO/0511).

### Randomization and Masking

After the baseline assessment, randomization of patients was conducted centrally at VU Amsterdam (Netherlands) by an independent researcher (EK) using automated Random Allocation Software (Department of Anaesthesia, Isfahan University of Medical Sciences). The allocation, stratified by country, was 1:1. Block randomization was used with variable block sizes, ranging from 2 to 14 allocations per block. In the United Kingdom, the process was outsourced to a local independent researcher for logistical reasons, which required the randomization outcome to be available immediately. Furthermore, stratification by recruitment location was applied at the national level for France (11 centers), Switzerland (3 centers), the Netherlands (3 centers), and Sweden (2 cities). Researchers and clinicians were blinded to the randomization scheme, but blinding of treatment allocation was impossible because both groups were evident to the therapists and patients.

### Sample Size Calculation

RCTs can determine whether a treatment effect is statistically significant and quantify its magnitude, but they do not necessarily capture the clinical relevance of the findings. Several approaches exist to assess clinical importance. In this study, we adopt the minimal clinically important difference from the patient perspective, as proposed by Cuijpers and colleagues [32], who identified a threshold of Cohen $d$=0.24. Given our noninferiority design, we applied a more conservative margin of $d$=0.20 for the primary outcome (PHQ-9). Under this criterion, bCBT is considered noninferior to TAU in clinical terms if the upper bound of the confidence interval does not exceed this predefined margin (1-sided tested). The required sample size was 1052 patients; to allow for dropouts, we aimed to enroll 1200 patients. This calculation was based on 90% power (type II error ≤0.10) and the requirement that the upper bound of the 2-sided 95% CI would not exceed the prespecified noninferiority margin.

### bCBT for MDDs

The bCBT protocols used in this study were based on ftfCBT protocols for MDD. The bCBT protocols integrated face-to-face sessions with structured internet-based treatment modules and ecological momentary assessments (EMAs) via smartphone to monitor daily mood, cognition, activity, social interaction, and



XSL•FO
RenderX

sleep. During these face-to-face sessions, the core principles of CBT were applied, alongside mandatory internet modules at all participating sites (Table 1). These mandatory modules focused on psychoeducation, cognitive restructuring, behavioral activation, and relapse prevention, as well as homework assignments. These internet-based modules were discussed in subsequent face-to-face sessions. The mandatory internet modules comprised 75% of the total number of modules. Sites also had the option to add extra modules (25%), such as emotional regulation or coping skills training, to facilitate adaptation to setting-specific practices, while maintaining protocol standardization across sites. At least one-third of the total number of sessions were to be held face-to-face, while another one-third were to be held online to ensure sufficient face-to-face and online modules.

**Table .** Overview of blended cognitive behavioral therapy applied within each country by platform, setting, duration, number of face-to-face and internet sessions, sequencing and mandatory internet modules (X), and additional modules.

| Country | Platform | Setting | Duration | Number of face-to-face and internet modules | Sequencing | Mandatory internet modules (X)[a] and additional modules |
|---|---|---|---|---|---|---|
| Germany | Moodbuster | PC[b] | 12 weeks | 6/10 | Alternate | • X<br>• PS[c]/PEX[d] |
| Poland | Moodbuster | PC | 8 weeks | 7/6 | Alternate | • X<br>• PS/PEX |
| Sweden | Iterapi | PC | 10 weeks | 4/6 | Alternate | • X<br>• Other |
| United Kingdom | Moodbuster | PC | 11 weeks | 6/5 | Alternate | • X<br>• PS/PEX |
| Spain | Smiling is Fun | PC/SMHC[e] | 10 weeks | 3/8 | 1-4-1-4-1 | • X<br>• PS/PE[f]/CS[g] |
| Denmark | NoDep | SMHC | 12 weeks | 12/7 | Alternate | • X<br>• other |
| France | Moodbuster | SMHC | 16 weeks | 8/8 | Alternate | • X<br>• PS/PEX |
| Netherlands | Moodbuster | SMHC | 20 weeks | 10/10 | Alternate | • X<br>• PS/PEX |
| Switzerland | Deprexis | SMHC | 18 weeks | 9/9 | Alternate | • X<br>• Other |

[a]X: mandatory modules.

[b]PC: primary care.

[c]PS: problem-solving.

[d]PEX: physical exercise.

[e]SMHC: specialized mental health care.

[f]PE: psychoeducation.

[g]CS: coping skills.

Table 1 summarizes, by country, the setting, duration, intensity, the alternation between face-to-face sessions and online modules, and additional modules for the bCBT group. As we followed routine care procedures, 5 countries used their own bCBT platforms: Denmark, the Netherlands, Spain, Sweden, and Switzerland. Four countries—France, Germany, Poland, and the United Kingdom—implemented a translated version of the Dutch Moodbuster platform for their patients because no bCBT platforms were available in these countries. All these platforms complied with General Data Protection Regulation requirements [33]. bCBT was delivered within primary care in 4 countries, while in Spain, it was provided both within primary care and outpatient specialized mental health care due to constraints within the original specialized mental health care setting. In the other 3 countries, it was delivered within specialized outpatient mental health care settings. bCBT began with a face-to-face session in all the countries. Within primary care, the treatment duration of bCBT ranged from 6 to 13 weeks, comprising 3-7 face-to-face sessions and 5-10 internet modules. Within specialized mental health care services, the treatment duration was longer, typically from 12 to 20 weeks, with 8-12 face-to-face sessions and 7-10 internet modules. All countries alternated face-to-face sessions with online modules, except for Spain (which used 2 face-to-face sessions, 4 internet modules,



XSL•FO
RenderX

and a final face-to-face session). The total number of ftfCBT sessions ranged from 3 to 6 in primary care settings and from 8 to 12 in specialized mental health care settings, and sessions were conducted in accordance with guidelines for treating moderate to severe MDD [34]. These face-to-face sessions lasted between 45 and 60 minutes each. The internet modules required 30-45 minutes of patients' time and needed 3-5 minutes daily to answer the EMA questions. The therapists used the platforms to monitor their patients' progress and to communicate securely with them. They were required to spend approximately 15 minutes each week per patient to review their online homework assignments. CBT therapists comprised licensed professionals and supervised psychology trainees; supervision by a licensed psychologist specializing in CBT applied only in primary care settings.

### Treatment as Usual

All participants in the control group received active treatment. TAU comprised routine care for patients with MDD in the treatment settings from which they were recruited (primary care or specialized mental health care). Six countries delivered ftfCBT within routine care (Denmark, France, the Netherlands, Poland, Switzerland, and the United Kingdom). In Sweden, Germany, and partly in Spain, TAU consisted of generic general practitioner care, which might have included watchful waiting, referral to a mental health specialist, psychotherapy (in Sweden and Germany), pharmacological treatment, or a combination of these. In the ftfCBT settings, therapists' professional qualifications were similar to those in the bCBT groups.

### Outcomes, Measures, and Time Points

All measures were self-reported and completed online (except for the MINI, which was assessed by independent interviewers blinded to the groups). Primary and secondary outcomes (including costs) were assessed at baseline and at the 3, 6, and 12-month follow-up periods. The MINI was assessed only at baseline and at the 12-month follow-up. We followed routine care RCT protocols in the participating countries, which resulted in primary care studies being shorter in duration and lower in intensity than those conducted in specialized mental health care settings. Posttreatment assessments meant that all participants could have completed their treatment 6 months after baseline. To enable an overall comparison, we therefore defined the 3-month assessment as a during-treatment measurement, the 6-month assessment as a posttreatment measurement, and the 12-month assessment as a follow-up measurement. The outcomes and time points are presented in Table 2. The reliability of the questionnaires is presented in Multimedia Appendix 2. Cronbach α ranged from 0.70 to 0.98, indicating good overall reliability.



XSL•FO
RenderX

**Table .** Overview of measures and time points.

| Variable | Measures | Baseline | 3 months | 6 months | 12 months |
|---|---|---|---|---|---|
| Questions asked to patients | | | | | |
| Demographics | N/A[a] | ✓ | N/A | N/A | N/A |
| Diagnostic interview | MINI[b] | ✓ | N/A | N/A | ✓ |
| Depressive symptoms | PHQ-9[c] and QIDS-SR16[d] | ✓ | ✓ | ✓ | ✓ |
| Quality of life | EQ-5D-5L | ✓ | ✓ | ✓ | ✓ |
| Societal and health care costs | TiC-P[e] | ✓ | ✓ | ✓ | ✓ |
| Treatment preference | N/A | ✓ | N/A | N/A | N/A |
| Patient expectancy and credibility | CEQ[f] | ✓ | N/A | N/A | N/A |
| Working alliance | WAI-SF-C[g] | N/A | ✓ | N/A | N/A |
| Technology alliance | WAI-TECH-SF[h] | N/A | ✓ | N/A | N/A |
| Patient satisfaction | CSQ[i] | N/A | ✓ | N/A | N/A |
| Satisfaction with the online program | SUS[j] | N/A | ✓ | N/A | N/A |
| Questions asked to therapists | | | | | |
| Working alliance | WAI-SF-Therapists | N/A | ✓ | N/A | N/A |

[a]N/A: not applicable.

[b]MINI: Mini International Neuropsychiatric Interview.

[c]PHQ-9: Patient Health Questionnaire-9.

[d]QIDS-SR 16: Quick Inventory of Depressive Symptomatology Self-Report-16.

[e]TiC-P: Trimbos/iMTA questionnaire for Costs associated with Psychiatric illness.

[f]CEQ: Credibility and Expectancy Questionnaire.

[g]WAI-SF-C: Working Alliance Inventory-Short Form–Client.

[h]WAI-TECH-SF: Working Alliance Inventory for Online Interventions-Short Form.

[i]CSQ: Client Satisfaction Questionnaire.

[j]SUS: System Usability Scale.

## Primary Outcome

The PHQ-9 was the primary outcome at posttreatment. It has demonstrated good psychometric properties in both primary and specialized care populations [30,35,36]. The total score ranges from 0 to 27, with 0‐4 indicating no to minimal depression, 5‐9 indicating mild depression, 10‐14 indicating moderate depression, 15‐19 indicating moderately severe depression, and 20‐27 indicating severe depression. Paper-based and online versions of the PHQ-9 yield comparable psychometric properties [37]. The PHQ-9 was also used to assess reliable deterioration. Participants were classified as having reliably deteriorated when the increase in PHQ-9 score exceeded the Reliable Change Index (RCI) threshold (|RCI|>1.96) [38]. See Multimedia Appendix 3 for a detailed description of the deterioration calculation.

## Secondary Outcomes

The Quick Inventory of Depressive Symptomatology Self-Report-16 (QIDS-SR 16) was used as a second measure to assess the level of depressive complaints [39]. The MINI [31] was used to evaluate MDD and comorbid anxiety disorders at baseline and at the 12-month follow-up to assess depression-free episodes (remission).

## Additional Outcomes

### *Overview*

Patient treatment expectations were measured with the Credibility and Expectancy Questionnaire (maximum score of 27 [40]) at baseline only. The remaining 4 outcomes were assessed during treatment, 3 months after the baseline assessment. Treatment satisfaction was assessed using the Client Satisfaction Questionnaire-8 (CSQ-8; range 8-32 [41]), and satisfaction with the online platforms was evaluated using the System Usability Scale (SUS) [42]. The therapeutic alliance between patients and therapists was measured using the short version of the Working Alliance Inventory, which comprises the Bond, Task, and Goal subscales, and a composite score (Working Alliance Inventory-Short Form [WAI-SF]) for patients [43] and therapists [44]. For descriptive purposes, patients' and



XSL•FO
RenderX

therapists' ratings were transformed into a 0-5 range for each component separately (as well as for the composite score). Patients' alliance with online platforms (experimental group only) was evaluated using the Working Alliance Inventory–Technology (Working Alliance Inventory for Online Interventions-Short Form [WAI-TECH-SF] [45], derived from the WAI-SF). Demographic variables such as age, sex, education, marital status, income, and employment, along with mental health treatment history, were assessed only at baseline. Patients' preference for bCBT or TAU was determined using a simple yes/no/no-preference question.

### *Cost Measures*

Costs were evaluated from the health care perspective, based on health care and patient costs, using the TiC-P (Trimbos/iMTA questionnaire for Costs associated with Psychiatric illness) questionnaire at baseline and at 3, 6, and 12 months [46]. To avoid cost discrepancies due to varying prices across countries, resource usage was assessed using standard Dutch costs [47]. For reference, the average 2017 exchange rate was €1=US $1.1297 (European Central Bank).

## Data Collection and Analyses

Each participating country collected data in accordance with a predefined data codebook, which was then uploaded to a centralized database (Research Department of GGZ inGeest Mental Health Care). The encrypted data, free of identifiable patient and therapist information, complied with General Data Protection Regulation regulations [33].

## Statistical Analyses

Analyses were based on the intention-to-treat (ITT) principle. We initially considered a per-protocol analysis, as ITT approaches in noninferiority trials may dilute between-group differences by including nonadherent participants, potentially biasing results toward noninferiority [48]. However, we ultimately adopted an ITT analysis because per-protocol can compromise the benefits of randomization and introduce selection bias. In addition, ITT better reflects routine clinical practice, where nonadherence is common, thereby enhancing generalizability to routine care settings. It also captures real-world effectiveness, aligns with health economic evaluation objectives, and accounts for variability in adherence and outcomes.

Descriptive statistics were used to assess baseline characteristics of the participants (Table 3). Multiple imputation was conducted using chained equations with predictive mean matching until the loss of efficiency was less than 5%, resulting in 20 complete datasets [49]. Linear mixed modeling [50] estimated the mean levels of the outcome variable for the TAU group and the intervention effects at 3 time points, using 3 time dummies (3-, 6-, and 12-month measurements) and 3 time dummy-by-group interaction terms; hypothesis tests for these interaction effects were conducted 1-sided, in line with the noninferiority framework. Alongside these fixed effects (for the main estimates), the model included random effects (random intercepts and time slopes for countries) and allowed for correlated residuals (for measurements within individuals). Missing observations for the CSQ and SUS were not imputed, as the answers to these questions were highly subjective. This reduces the risk of bias due to incorrect imputation [51]; AD use at the 3-, 6-, and 12-month assessments was also not imputed. Estimated intervention effects (EE) were presented on the corresponding scales for the 4 outcomes. Cohen *d* and 95% CIs [52] were used to quantify between-group treatment effect sizes for both the PHQ-9 and the QIDS. Effect sizes were classified as very small ($d<0.2$), small (0.2-0.5), medium (0.5-0.8) or large (>0.8). MINI-D (Depression) and MINI-A (Anxiety) were presented as log-odds ratios using the Hasselblad and Hedges logit method [53]. A "leave-one-out" sensitivity analysis assessed the robustness of the overall effect by systematically excluding one country at a time and recalculating the effect size for the primary outcome [54]. To assess the impact of baseline AD use on treatment outcomes during treatment and follow-ups, we extended the linear mixed modeling by adding a third factor (AD use: yes/no). This allowed estimation of marginal mean PHQ-9 scores at different time points for the bCBT and TAU groups within strata of AD use, as well as effect sizes. We calculated a reliable deterioration in the PHQ-9 [38] during treatment (3 months), posttreatment (6 months), and at the 12-month follow-up point. The nonimputed dataset assessed the number of patients who deteriorated, while the log ratios and corresponding effect sizes were estimated using generalized linear (logistic) mixed models. Stata (version 14.2; StataCorp LLC) was used to conduct the analyses.



XSL•FO
RenderX

**Table .** Baseline sociodemographic patient characteristics.

| Characteristics | bCBT[a] | TAU[b] | Total |
|---|---|---|---|
| Patients, n | 423 | 412 | 835 |
| Age (years), mean (SD) | 39.52 (13.27) | 38.84 (13.09) | 39.17 (13.18) |
| Sex (female), n % | 282 (66.67) | 281 (68.20) | 563 (67.42) |
| Marital status, n (%) | | | |
| Single | 142 (33.57) | 132 (32.04) | 274 (32.81) |
| Divorced | 55 (13.00) | 39 (9.47) | 94 (11.26) |
| Widowed | 2 (0.47) | 4 (0.97) | 6 (0.72) |
| Living together | 87 (20.57) | 103 (25.00) | 190 (22.75) |
| Married | 137 (32.39) | 134 (32.52) | 271 (32.46) |
| Education, n (%)[c] | | | |
| Low | 54 (12.77) | 60 (14.56) | 114 (13.65) |
| Middle | 156 (36.88) | 145 (35.19) | 301 (36.05) |
| High | 213 (50.35) | 207 (50.24) | 420 (50.30) |
| Employed, n (%) | N=418 | N=409 | N=827 |
| Yes | 248 (59.33) | 263 (64.30) | 511 (61.79) |
| No | 170 (40.67) | 146 (35.70) | 316 (38.21) |
| MINI[d], n (%) | | | |
| Major Depressive Disorder | 423 (100) | 412 (100) | 835 (100) |
| Anxiety Disorder | 221 (52.25) | 199 (48.30) | 420 (50.30) |
| PHQ-9[e], mean (SD) | 15.36 (4.82) | 15.21 (4.53) | 15.28 (4.67) |
| QIDS[f], mean (SD) | 14.60 (4.40) | 14.65 (4.23) | 14.63 (4.31) |
| Antidepressant use, n/N (%) | 168/417 (40.29) | 184/406 (45.32) | 352/823 (42.77) |
| Treatment preference, n (%) | | | |
| bCBT | 241 (56.97) | 230 (55.83) | 471 (56.41) |
| TAU | 73 (17.26) | 64 (15.53) | 137 (16.41) |
| No preference | 109 (25.77) | 118 (28.64) | 227 (27.18) |
| CEQ_cred, mean (SD)[g] | 19.45 (4.73) | 16.97 (6.13)[h] | 18.32 (5.55) |
| CEQ_exp, mean (SD)[g] | 17.57 (4.96) | 15.45 (6.09)[h] | 16.60 (5.61) |
| UTILNL[i], mean (SD) | 0.58 (0.23) | 0.57 (0.24) | 0.57 (0.23) |
| Health care costs (€, TiC-P[j], 4 weeks), mean (SE)[j] | 1172 (185) | 1338 (183) | 1255 (12.74) |

[a]bCBT: blended cognitive behavioral therapy.

[b]TAU: treatment as usual.

[c]Data collected were regarding what would be considered low, middle, and high levels of education in each setting.

[d]MINI: Major Depressive Disorder, Anxiety Disorder.

[e]PHQ-9: Patient Health Questionnaire-9.

[f]QIDS: Quick Inventory of Depressive Symptomatology Self-Report.

[g]CEQ cred and CEO exp: Credibility and Expectancy Questionnaire [40].

[h]$P$ <.001.

[i]UTILNL: scores, Netherlands [55].

[j]TiC-P: Trimbos/iMTA questionnaire for Costs associated with Psychiatric illness [46].



XSL•FO
RenderX

### Cost-Effectiveness Analyses

Cost-effectiveness analyses were conducted from the health care perspective, as this perspective best aligned with our hypothesis. This perspective is also recommended by the National Institute for Health and Care Excellence in the United Kingdom [56]. Health-related quality of life for the cost-effectiveness analyses was measured using the EQ-5D-5L [57]. EQ-5D-5L health states were converted into utility scores using the Dutch tariff [58], and quality-adjusted life years (QALYs) were calculated by multiplying utilities by patient time spent in each health state, with linear interpolation applied between transitions.

Within the multiple imputation procedure, predictive mean matching was used to account for the skewed distribution of cost data. Mixed-effects regression models with repeated measurements clustered within participants were used to estimate differences in overall PHQ-9 scores and the number of MDD cases prevented over the 12-month follow-up period. Differences in QALYs and costs were analyzed using linear regression models.

Incremental cost-effectiveness ratios (ICERs) were calculated by dividing the difference in costs between the 2 conditions by the corresponding difference in effects. Statistical uncertainty surrounding the ICERs was assessed using bias-corrected and accelerated bootstrapping with 5000 replications and visualized in cost-effectiveness planes. Cost-effectiveness acceptability curves were constructed to estimate the probability that bCBT was cost-effective compared with TAU across a range of willingness-to-pay (WTP) thresholds. The WTP threshold represents the maximum amount society is willing to pay for 1 additional unit of health gain. No established WTP threshold exists for improvements in PHQ-9 scores. For QALYs, however, a maximum WTP threshold of €50,000 per QALY gained was applied, in accordance with the Dutch Health Care Institute's recommendation for interventions targeting conditions with an estimated disease burden of 0.48. Finally, a sensitivity analysis was conducted in which differences in costs and effects were adjusted for potential confounders, including baseline health care costs, age, sex, and treatment setting (primary vs secondary care) [55,56]. Analyses were conducted in Stata (version 14.2), and reporting adhered to the CHEERS (Consolidated Health Economic Evaluation Reporting Standards) [59].

## Results

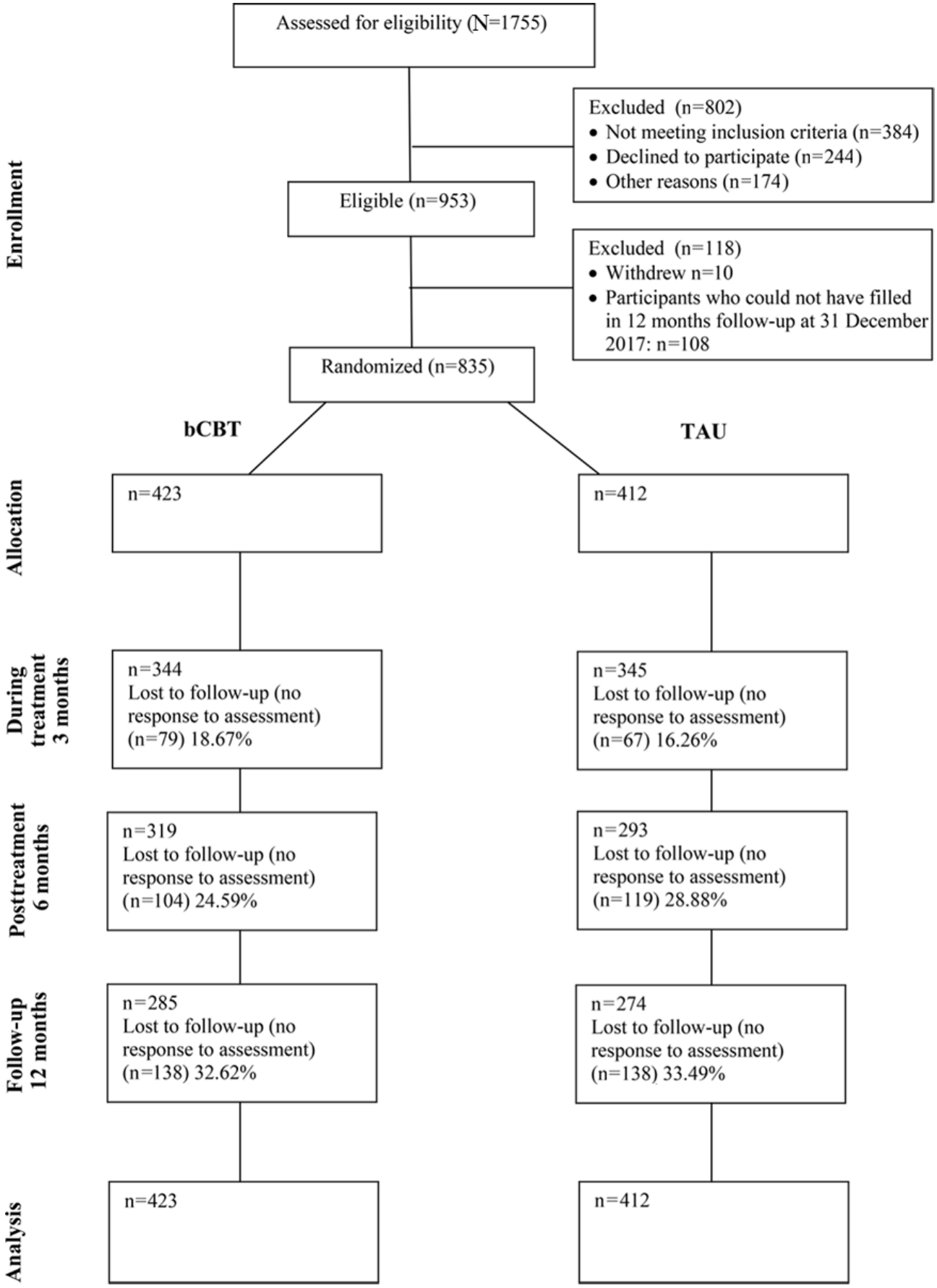
### Overview

Figure 1 illustrates the study's participant flow. Of 1755 participants assessed for eligibility, 953 started treatments, of whom 118 were excluded from our analysis. Ten participants withdrew before randomization, and for 108, the 12-month assessment would be beyond December 31, 2017 (the funder-determined time frame). Thus, the ITT sample included 835 participants, with 423 in bCBT and 412 in TAU. The total study dropout rates for the sample were 17.5% at 3 months after baseline assessment, 26.7% after 6 months, and 33% after 12 months.



XSL•FO
RenderX

**Figure 1.** Flow diagram of participants through the trial based on intention-to-treat (CONSORT [Consolidated Standards of Reporting Trials]). bCBT: blended cognitive behavioral therapy; TAU: treatment as usual.






XSL•FO
RenderX

Table 3 summarizes the participants' baseline sociodemographic, clinical, and cost profiles. Multimedia Appendix 4 presents participant counts and baseline characteristics for each country. The mean age of the total group was 39.17 (SD 13.18) years, with 67% female and half of the sample having a high level of education. The employment rate was 61.80%. The mean PHQ-9 score for the total group was 15.28 (SD 4.7), indicating moderately severe symptoms. Over half (420/835, 50.30%) of the participants had a comorbid anxiety disorder. At baseline, on average, 42.77% (352/823) of the participants used ADs, with no significant differences between the 2 groups. This percentage remained stable over time, averaging 42.65% (351/823) at 3 months, 38.40% (316/823) at 6 months, and 40.70% (335/823) at 12 months, with nonsignificant differences between the groups. At prerandomization, 56.41% (471/835) of the participants preferred bCBT, 16.41% (137/835) preferred TAU, and the remainder reported no preference. The bCBT group scored significantly higher on the credibility and expectancy scales for treatment than the TAU group, indicating moderate to high levels on both measures. Mean health care costs per patient 4 weeks prebaseline were €1172 (SE €185) for the bCBT group and €1338 (SE €183) for the TAU group.

## Primary Outcomes

Table 4 and Figure 2 show that, for the primary outcome (PHQ-9), bCBT was noninferior to TAU at 3 (Figure 2A), 6 (Figure 2B), and 12 (Figure 2C) months, as all upper bounds of the confidence intervals lay within the noninferiority margin of ≥0.20. A "leave-one-out" sensitivity analysis also showed noninferiority of the primary outcome at the 3- and 6-month assessments. However, at 12 months, the "leave-one-out" analysis revealed that excluding Germany yields inferior results. This indicates that the overall effect estimate is sensitive to this omission and suggests potential heterogeneity across settings. Together, these findings highlight uncertainty in the 12-month results and suggest that the observed effects may be context-dependent.

**Table .** Estimated intervention effects and effect sizes on both the primary outcome (PHQ-9)[a] and secondary outcomes (QIDS[b], MINI[c] Depression, and MINI Anxiety), based upon both 1-stage models with random time slopes based upon multiple imputed data (N=835).

| Outcomes | 3 months | 6 months | 12 months |
|---|---|---|---|
| Primary outcome: PHQ-9 | | | |
| EE[d] (SE) | −1.21 (0.41); *z*=−2.939; *P*=.003 | −1.00 (0.41); *z*=−2.411; *P*=.02 | −0.05 (0.42); *z*=−0.124; *P*=.90 |
| ES[e] (CI) | *d*=−0.26 (−0.43 to −0.09) | *d*=−0.21 (−0.39 to −0.04) | *d*=−0.01 (−0.19 to 0.17) |
| Secondary outcomes | | | |
| QIDS-16 | | | |
| EE (SE) | −1.15 (0.46); *z*=−2.489; *P*=.01 | −0.33 (0.39); *z*=−0.831; *P*=.41 | 0.14 (0.42); *z*=0.340; *P*=.73 |
| ES (CI) | *d*=−0.27 (−0.48 to −0.06) | *d*=−0.08 (−0.25 to 0.10) | *d*=0.03 (−0.16 to 0.22) |
| MINI-Depression | | | |
| OR[f] (CI) | N/A[g] | N/A | 0.67 (0.45 to 0.99) |
| MINI-Anxiety | | | |
| OR (CI) | N/A | N/A | 0.78 (0.46 to 1.32) |

[a]PHQ-9: Patient Health Questionnaire-9.

[b]QIDS: Quick Inventory of Depressive Symptomatology.

[c]MINI: Mini International Neuropsychiatric Interview.

[d]EE: estimated intervention effects.

[e]ES: effect sizes.

[f]OR: odds ratio.

[g]N/A: not applicable.



XSL•FO
RenderX

**Figure 2.** Leave-one-out meta-analysis (sensitivity analysis). Primary outcome PHQ-9: effect sizes and confidence intervals at (A) 3 months (during treatment), (B) 6 months, and (C) 12 months (N=835). PHQ-9: Patient Health Questionnaire-9.



XSL•FO
RenderX

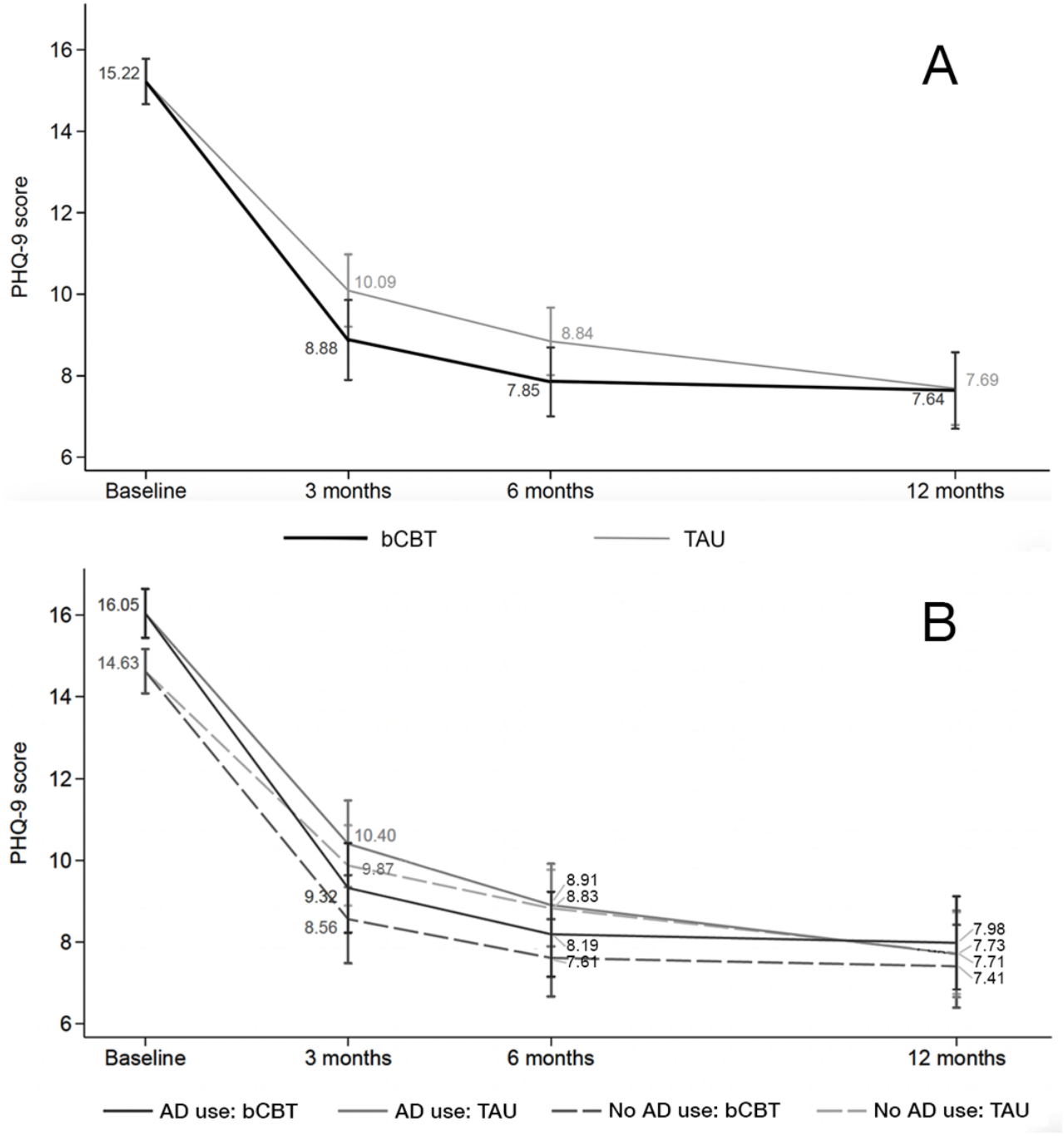
## Secondary Outcomes

### *Overview*

Noninferiority of bCBT is also shown in Table 4 for the secondary QIDS outcome at 3 months during treatment ($d$=−0.27, 95% CI −0.48 to −0.06) and at posttreatment (6 months) ($d$=−0.08, 95% CI −0.25 to 0.10) but not at 12 months' follow-up (95% CI −0.16 to 0.22). The bCBT group also had a lower risk of MDD than the TAU group at the 12-month follow-up (odds ratio [OR] 0.67, 95% CI 0.45-0.99) for anxiety disorders; this difference was not significant (OR 0.78, 95% CI 0.46-1.32).

As shown in Table 4 and Figure 3A, both groups exhibited significant within-group changes in the primary outcome, with a rapid decrease leading up to posttreatment at 6 months. Differences were small but significant for bCBT over TAU during treatment (3 months, $d$=−0.26, 95% CI −0.43 to −0.09; $P$=.003) and at 6 months posttreatment ($d$=−0.21, 95% CI −0.39 to −0.04; $P$=.016), with a nonsignificant difference between the groups at the 12-month follow-up.

**Figure 3.** Estimated marginal means PHQ-9 scores at various time points for the bCBT and TAU group (A) and for strata of AD use (B). AD: antidepressant; bCBT: blended cognitive behavioral therapy; CBT: cognitive behavioral therapy; PHQ-9: Patient Health Questionnaire-9; TAU: treatment as usual.




A subgroup analysis (Table 5 and Figure 3B) was conducted to estimate the impact of baseline AD versus non-AD use on the primary outcome (PHQ-9) in bCBT versus TAU. At baseline, the AD groups had a 1.42-point higher PHQ-9 score than the



XSL•FO
RenderX

non-AD user groups, indicating that this group experienced more severe depressive symptoms. AD patients exhibited a greater reduction in depressive symptoms than non-AD users, most likely because their higher baseline scores left more room for improvement (Figure 3). When comparing the impact of AD use and non-AD use on the primary outcome across the 2 groups, no significant differences were found at any of the 3 time points (Table 5, AD vs no AD at baseline, bCBT compared with TAU).

**Table .** Estimated marginal mean PHQ-9[a] scores at various time points for the bCBT[b] and TAU[c] groups and by strata of AD[d] use.

| | Difference in EMM[e] (95% CI) | Difference in Cohen *d* (95% CI) | *P* value |
|---|---|---|---|
| No AD at baseline bCBT-TAU | | | |
| 3 months | −1.31 (−2.35 to −0.28) | −0.28 (−0.50 to −0.06) | .01 |
| 6 months | −1.21 (−2.28 to −0.15) | −0.26 (−0.49 to −0.03) | .02 |
| 12 months | −0.32 (−1.41 to 0.77) | −0.07 (−0.30 to 0.16) | .56 |
| AD at baseline bCBT-TAU | | | |
| 3 months | −1.08 (−2.26 to 0.10) | −0.23 (−0.48 to 0.02) | .07 |
| 6 months | −0.71 (−1.95 to 0.52) | −0.15 (−0.42 to 0.11) | .26 |
| 12 months | 0.27 (−0.97 to 1.52) | 0.06 (−0.21 to 0.32) | .66 |
| AD versus no AD at baseline bCBT-TAU | | | |
| 3 months | 0.23 (−1.26 to 1.73) | 0.05 (−0.27 to 0.37) | .76 |
| 6 months | 0.50 (−1.12 to 2.12) | 0.11 (−0.24 to 0.45) | .54 |
| 12 months | 0.59 (−1.03 to 2.22) | 0.13 (−0.22 to 0.47) | .47 |

[a]PHQ-9: Patient Health Questionnaire-9.

[b]bCBT: blended cognitive behavioral therapy.

[c]TAU: treatment as usual.

[d]AD: antidepressant.

[e]EMM: estimated marginal means.

Using the nonimputed dataset for the entire sample, the overall percentage of patients who deteriorated was low: 4.8% (33/688) at the 3-month assessment, 4.6% (28/609) at the 6-month assessment, and 4.1% (23/561) at the 12-month assessment. There were no significant differences in EE on reliable deterioration between the groups at all 3 time points (OR and Cohen *d*). However, the estimated multilevel logistic model for deterioration yielded unreliable estimates, most likely due to the small number of patients who deteriorated.

### *Acceptance and Working Alliance*

Table 6 presents results for several acceptance and working alliance measures. Patients in the bCBT group scored significantly higher on treatment satisfaction (CSQ-8) than those in the TAU group, with a mean of 25.44 (SD 4.93) at the 3-month assessment. This score falls between “somewhat satisfied” (n=24) and “very satisfied” (n=32), whereas the TAU group scored below the 24 threshold (mean 21.34, SD 5.94). Only the bCBT group (patients) rated the technological platforms’ usability, with a mean SUS score of 40.16 (SD 7.35), below the average of 68. In contrast, the technical alliance (WAI-TECH-SF-C48), also rated only by the bCBT group (patients), was rated more positively, with a mean of 57.12 (SD 16.87) on a scale of 12 - 84. Higher scores indicate a better quality of the technical alliance.



XSL•FO
RenderX

**Table .** Mean scores for user satisfaction and working and technical alliances by bCBT[a] and TAU[b].

| | bCBT | TAU |
|---|---|---|
| CSQ[c], N (mean, SD) | 203 (25.51, 4.71) | 194 (21.34, 5.94)[d] |
| SUS[e], N (mean, SD) | 324 (40.16, 7.35) | N/A[f] |
| WAI-SR-C[g], composite, N (mean, SD) | 265 (3.97, 0.77) | 240 (3.41, 0.99)[d] |
| WAI-SR-C task, N (mean, SD) | 265 (3.86, 0.84) | 240 (3.24, 1.13)[d] |
| WAI-SR-C bond, N (mean, SD) | 265 (4.19, 0.83) | 240 (3.78, 1.03)[d] |
| WAI-SR-C goal, N (mean, SD) | 265 (3.86, 0.84) | 240 (3.21, 1.10)[d] |
| WAI-TECH-SF-C[h], N (mean, SD) | 297 (57.12, 16.87) | N/A |
| WAI-SR-Therap[i] composite, N (mean, SD) | 243 (3.84, 0.70) | 216 (4.09, 0.64)[d] |
| WAI-SR-Therap task, N (mean, SD) | 243 (3.54, 0.87) | 216 (3.88, 0.79)[d] |
| WAI-SR-Therap bond, N (mean, SD) | 243 (4.33, 0.63) | 216 (4.41, 0.60) |
| WAI-SR-Therap goal, N (mean, SD) | 243 (3.65, 0.90) | 216 (3.99, 0.78)[d] |

[a]bCBT: blended cognitive behavioral therapy.

[b]TAU: treatment as usual.

[c]CSQ: Client Satisfaction Questionnaire.

[d]*P*<.001.

[e]SUS: System Usability Scale.

[f]N/A: not applicable.

[g]WAI-SR-C: Working Alliance Inventory-Short Revised–Client.

[h]WAI-TECH-SF-C: Working Alliance Inventory for Online Interventions-Short Form–Client.

[i]WAI-SR-Therap: Working Alliance Inventory-Short Revised–Therapists.

Patients in bCBT and TAU rated the working alliance as good, with mean overall composite scores of 3.97 (SD 0.77) in the bCBT group and 3.41 (SD 0.99) in the TAU group, out of 5. However, patients in the bCBT group rated the therapeutic alliance (WAI-SR-C) significantly higher (*P*<.001) than those in the TAU group, both for the composite score and for the task, bond, and goal components. The therapists also rated the working alliance as good, but those in the TAU group rated it significantly higher (mean 4.09, SD 0.64; *P*<.001) than those in the bCBT group (mean 3.84, SD 0.70). They also scored significantly higher on the task (*P*<.001) and goal (*P*<.001) components of the WAI-SR-T in TAU compared with bCBT but not on the bond scale.

## Costs

Table 7 presents several disaggregated health care cost categories and total health care costs per patient for bCBT and TAU at the 12-month assessment. These costs did not differ significantly between the 2 groups, except for home care costs, which were significantly higher for TAU than for bCBT. However, the average total health care costs per patient were slightly higher for the bCBT group than for the TAU group (estimated mean difference per patient €315, 95% CI −582 to 1163). This difference was primarily driven by the estimated bCBT intervention costs.



XSL•FO
RenderX

**Table .** Estimated mean (SE) costs (euros) per patient for bCBT[a] and TAU[b], and corresponding differences (95% CI) at 12-month follow-up point (N=835).

| Cost category | bCBT (€/SE) | TAU (€/SE) | Difference (95% CI) |
|---|---|---|---|
| Intervention costs | 274 (6) | 0 (0) | N/A[c] |
| Primary care costs | 440 (31) | 525 (49) | −85 (−213 to 7) |
| Mental health care costs | 1517 (159) | 1458 (142) | 59 (−318 to 455) |
| Complementary care costs | 74 (13) | 55 (13) | 18 (−16 to 48) |
| Secondary care costs | 1131 (230) | 745 (129) | 386 (−25 to 900) |
| Home care costs | 687 (96) | 1029 (200) | −342 (−867 to −8) |
| Mental health medication costs | 35 (3) | 33 (3) | 3 (−4 to 9) |
| Total health care costs | 4159 (345) | 3845 (331) | 315 (−582 to 1163) |

[a]bCBT: blended cognitive behavioral therapy.

[b]TAU: treatment as usual.

[c]Not applicable.

## Cost-Effectiveness Analyses

Table 8 presents the cost-effectiveness analysis results at the 12-month follow-up from a health care perspective. The mean overall change in PHQ-9 score over the 12-month period was small but significantly higher for bCBT than for TAU. Within the main analysis, the ICER for the PHQ-9 score was €643 per point of improvement, meaning that to gain a single point of improvement in the PHQ-9 score, an investment of €643 is required for bCBT compared with TAU.



XSL•FO
RenderX

**Table .** Multiple imputed, pooled cost-effectiveness outcomes for bCBT[a] compared with TAU[b] (euros) at the 12-month follow-up point[c].

| Outcome | ΔC[d] (95% CI) (€) | ΔE[e] (95% CI) | ICER[f] (€) | CE[g] plane | | | |
|---|---|---|---|---|---|---|---|
| | | | | NE[h] | SE[i] | SW[j] | NW[k] |
| Main analysis | | | | | | | |
| PHQ-9[l] | 315 (−533 to 1168) | 0.49 (0.01 to 0.97) | 643 | 73% | 25% | 0% | 2% |
| MINI[m] | 315 (−533 to 1168) | 0.07 (0.01 to 0.14) | 4381 | 74% | 25% | 0% | 1% |
| QALY[n] | 315 (−533 to 1168) | 0.01 (−0.02 to 0.04) | 36,729 | 51% | 21% | 4% | 24% |
| Adjusted analysis | | | | | | | |
| PHQ-9 | 421 (−319 to 1161) | 0.49 (0.02 to 0.97) | 852 | 82% | 16% | 0% | 2% |
| MINI | 421 (−319 to 1161) | 0.07 (0.01 to 0.14) | 5653 | 83% | 16% | 0% | 1% |
| QALY | 421 (−319 to 1161) | 0.01 (−0.02 to 0.04) | 45,210 | 60% | 13% | 2% | 25% |

[a]bCBT: blended cognitive behavioral therapy.

[b]TAU: treatment as usual.

[c]Differences in PHQ-9 score and MINI were multiplied by −1 in order to ensure that the CE planes and cost-effectiveness acceptability curves were interpretable. Thus, differences in PHQ-9 score should be interpreted as additional points of improvement in bCBT compared with TAU and differences in MINI as cases of MDD prevented within bCBT compared with TAU.

[d]ΔC: The difference in costs between blended care and usual care.

[e]ΔE: difference in effects between bCBT and TAU.

[f]CER: incremental cost-effectiveness ratio.

[g]CE: cost-effectiveness.

[h]NE: northeast.

[i]SE: southeast.

[j]SW: southwest.

[k]NW: northwest.

[l]PHQ-9: Patient Health Questionnaire-9.

[m]MINI: Mini International Neuropsychiatric Interview.

[n]QALY: quality-adjusted life-years (based on EQ-5D-5L).

The cost-effectiveness plane shows that in 73% of the bootstrapped simulations, bCBT was both more effective and more expensive (Table 7: northeast quadrant of the cost-effectiveness plane). Figure 4 present the cost-effectiveness planes, plotting differences in costs against differences in effects (PHQ-9 (Figure 4A), presence of MDD (Figure 5A), and QALYs (Figure 6A) between the 2 conditions combined with the cost-effectiveness acceptability curves (Figure 4B, Figure 5B, and Figure 6B) representing the probability that bCBT is cost-effective compared with TAU across a range of WTP thresholds (WTP for PHQ-9, presence of MDD, and QALYs). As no universally accepted thresholds exist for these outcomes (except for QALYs), WTP values were varied from €0 to €50,000 to reflect a broad range of decision-making scenarios.



XSL•FO
RenderX

**Figure 4.** Cost-effectiveness plane of incremental costs versus change in PHQ-9 score over 12 months (A) across willingness-to-pay thresholds (B). bCBT: blended cognitive behavioral therapy; PHQ-9: Patient Health Questionnaire-9; TAU: treatment as usual.

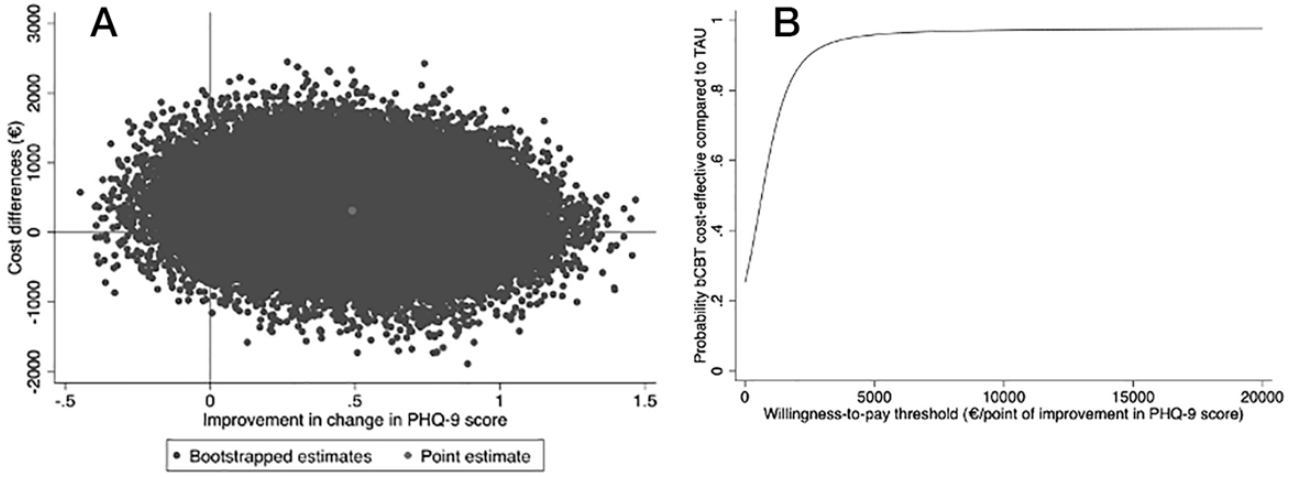



**Figure 5.** Cost-effectiveness plane of incremental costs versus change in presence of MDD (A) (MINI, 51) and across willingness-to-pay thresholds over 12 months (B). bCBT: blended cognitive behavioral therapy; MDD: major depressive disorder; MINI Mini International Neuropsychiatric Interview; TAU: treatment as usual.

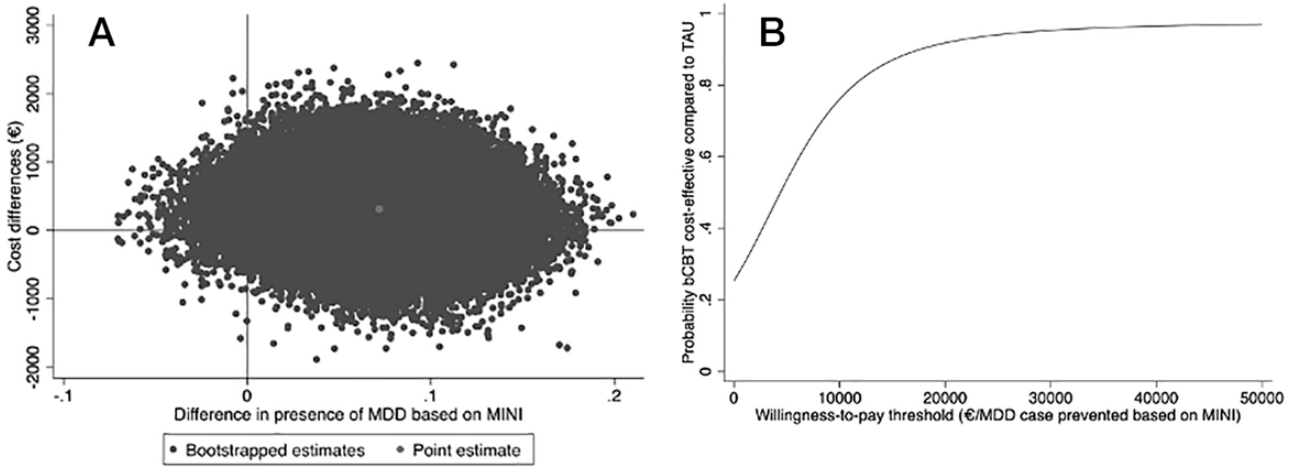
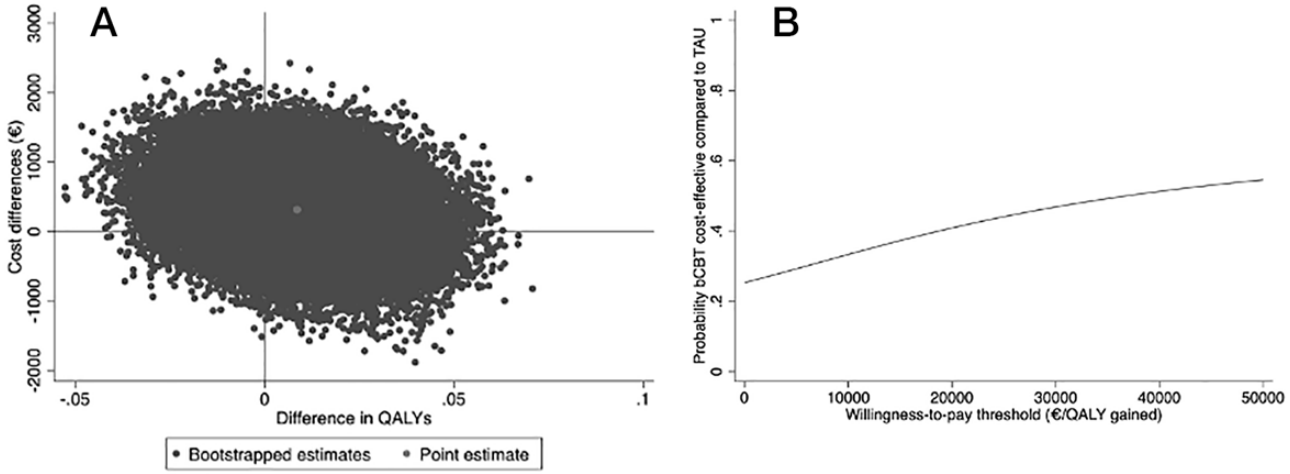



**Figure 6.** Cost-effectiveness plane of incremental costs versus change in QALYs (A) and acceptability curve across 12 months (B). bCBT: blended cognitive behavioral therapy; QALYs: quality-adjusted life-years; TAU: treatment as usual.




Cost-effectiveness planes illustrate the uncertainty around the ICER across the 4 quadrants. In Figure 4, 73% of bootstrap replications fall in the northeast quadrant, indicating that bCBT is not only more effective than TAU but also more costly (PHQ-9). It also shows that, for a 1-point improvement on the PHQ-9 and a WTP of €0, the probability of cost-effectiveness was low (0.25), rising to 0.95 at a WTP of €3,800.

A similar pattern was observed for preventing an MDD diagnosis at follow-up (12 months; Table 7). The ICER for preventing 1 MDD case was €4381, indicating that this amount



XSL•FO
RenderX

would need to be invested in bCBT rather than TAU to prevent an MDD diagnosis. As shown in Figure 5, 74% of the cost-effectiveness pairs fall within the northeast quadrant (more effective and more expensive), while the remaining 26% are in the southeast quadrant (more effective and less expensive). Figure 5 also shows that the probability of cost-effectiveness ranges from 0.25 at a WTP of €0 to 0.76 at a WTP of €10,000 for every MDD episode prevented.

For QALYs, the difference between bCBT and TAU was small and not statistically significant (Table 7), with considerable uncertainty indicated by the spread of bootstrap replications across all quadrants. The ICER indicates a cost of €36,729 per QALY gained. At a €50,000 threshold, the probability that bCBT is cost-effective is 0.55 (Figure 6). In the adjusted analysis, the cost difference between bCBT and TAU increased to €421 (Table 7), but this difference was not statistically significant, and the overall results were consistent with the main analysis.

## Discussion

### Principal Findings

Our hypotheses were that bCBT for patients with MDD would be clinically noninferior to TAU and more cost-effective from a health care perspective. The results showed that while bCBT was noninferior to TAU, it was not cost-effective. The first part of our hypothesis was confirmed, as all the differential effects observed for the primary outcome (PHQ-9) in the comparison between bCBT and TAU fell within the conservative noninferiority margin (Cohen *d*=0.20). Depressive symptoms showed a significant decline up to the posttreatment assessment, and this improvement was maintained at the 12-month follow-up. The posttreatment effects for both groups also aligned well with the benchmark (*d*=1.19 [60]) for the treatment of depression with CBT in routine care settings. Additionally, the long-term effects observed in both groups are comparable with those found in meta-analyses of CBT for depression [61,62]. The bCBT group also had a lower risk of experiencing an MDD episode than the TAU group at the 12-month follow-up (OR 0.67, 95% CI 0.45-0.99). AD use was evenly distributed between the 2 groups (approximately 40%) from baseline to the 12-month follow-up, with no significant differences in proportions. Subgroup analyses indicate that participants with and without baseline AD use benefited equally from treatment. These findings are comparable with those of a meta-analysis by Cuijpers and colleagues [63], which reported a similar percentage of AD use in psychotherapeutic RCTs and found that the effects of psychotherapy for adult depression were not concurrently mediated by pharmacotherapy use. Adverse treatment effects were limited, as indicated by the reliable deterioration rates (less than 5%) for the overall group up to the 12-month follow-up. These rates compare favorably with those found in a meta-analysis of deterioration rates within psychotherapies for depression [64].

Patients' acceptance of treatment, in terms of both the credibility and expectancy of therapy, was significantly higher for bCBT than for TAU, as was patients' satisfaction with treatment. This may be partly explained by the higher baseline preference for the bCBT group, which, in turn, was influenced by patients' willingness to participate in the study. An additional explanation for these higher satisfaction ratings may be that the internet-based components of bCBT better support patients' self-management than traditional face-to-face psychotherapy, as patients are less reliant on therapists only during their treatment. This may also partly explain why, although patients and therapists valued the working alliance as strong in both conditions, patients in the bCBT group gave higher ratings than those in the TAU group. Conversely, therapists rated the therapeutic alliance significantly higher in TAU than in bCBT, except on the bond subscale. This difference may reflect therapists' preference for face-to-face contact, including access to nonverbal cues, as well as established professional norms [65]. Nevertheless, the observed difference was relatively small, and other studies have not found such disparities [16]. Overall, these findings suggest that a strong working alliance—encompassing both human and technological components—can be established between patients and therapists in bCBT, although many participants in our study were relatively new to this mode of treatment delivery. The high alliance ratings reported by patients in both conditions are consistent with findings from the meta-analytic review by Flückiger et al [66]. For a more detailed analysis of the association between working alliance and treatment outcomes for E-COMPARED, see the study by Doukani et al [67].

Only a few RCTs have directly evaluated the clinical effectiveness of bCBT for MDD compared with active face-to-face treatment controls, limiting direct comparisons with our study. These trials recruited outpatients with MDD, compared bCBT with ftfCBT or face-to-face psychotherapy, and had broadly comparable treatment durations and intensities (12 - 20 weeks). Their findings are generally consistent with ours. For example, Berger et al [68] reported that CBT supplemented with internet-based modules was superior to ftfCBT at 12 weeks posttreatment (moderate effect) [68]. Similarly, Thase et al [69] demonstrated the noninferiority of integrated bCBT compared with ftfCBT 4 months after baseline assessment, whereas Kooistra et al [23] found no significant differences in depressive symptoms between the 2 treatment conditions. Berger et al [68] and Kooistra et al [23] also reported similarly low deterioration rates, good therapeutic alliance, and high treatment acceptability, with no significant between-group differences. These findings should, however, be interpreted cautiously because the available trials were relatively small and therefore provided limited certainty regarding treatment effects.

Since completion of our study, Schaeuffele et al [70] reported the results of the large pragmatic PsyTOM trial conducted in Germany, which compared a transdiagnostic blended psychotherapy intervention with routine face-to-face psychotherapy for patients with a broad range of mental health disorders. Based on a superiority design, the authors found no evidence that blended treatment resulted in greater clinical improvement or fewer face-to-face sessions up to 6 months after randomization [70]. Although these findings differ from ours, direct comparisons should be made cautiously because the interventions and study designs differed substantially. Whereas E-COMPARED evaluated a structured, integrated bCBT



XSL•FO
RenderX

protocol specifically designed for patients with MDD, PsyTOM evaluated a highly flexible transdiagnostic blended intervention delivered across multiple therapeutic orientations. In addition, PsyTOM included patients with a wide range of mental health disorders, did not restrict inclusion to formally diagnosed disorders, and allowed therapists considerable flexibility in determining both treatment duration and the use of digital components. The PsyTOM authors therefore suggested that the absence of superiority may reflect the flexibility and heterogeneity of the intervention rather than the ineffectiveness of blended treatment itself.

From a health care perspective, bCBT was associated with nonsignificantly higher health care costs than TAU over the 12-month follow-up period. Although differences in clinical outcomes were modest, they consistently favored bCBT, with slightly greater improvements in PHQ-9 scores and fewer MDD episodes. Consequently, whether bCBT can be considered cost-effective depends on how much policymakers are willing to pay for these additional health gains. Evidence on the cost-effectiveness of bCBT compared with active face-to-face treatment remains scarce. This is not unexpected, as relatively few pragmatic trials have evaluated blended care under routine clinical conditions, and economic evaluations have only occasionally been conducted alongside these trials. The few available studies have also reported inconsistent findings, with some reporting lower health care costs primarily because blended care reduced therapist time, whereas others found no significant cost differences. These inconsistencies likely reflect differences in study design, intervention characteristics, health care settings, cost perspectives, and economic evaluation methods rather than true differences in the cost-effectiveness of bCBT [23,69].

## Strengths and Limitations

This study is the first large-scale noninferiority trial to date, examining the clinical effectiveness and cost-effectiveness of bCBT compared with face-to-face TAU, predominantly ftfCBT. Noninferiority trials within routine mental health care remain scarce because of the large sample sizes required (small noninferiority margins necessitate high precision), the need for evidence-based treatment comparators, and evaluations under real-world conditions [71]. We addressed this gap by achieving a substantial sample size, using active comparators, setting a clear noninferiority margin, and conducting our RCT within routine care practices. Dropout rates in the total sample ranged from 17.5% at 3 months to 33% at the 12-month follow-up. The proportion of participants lost at 3 months (17.5%) was comparable with the average dropout reported in randomized trials of individual psychotherapy for depression in routine care [72], whereas the higher attrition at 12 months is consistent with the challenges of maintaining participant engagement during long-term follow-up. Despite this, dropout warrants careful consideration in future studies, as it may affect the validity of the findings and may indicate challenges with treatment adherence. The multifaceted nature of our study, which comprised diverse settings and technical platforms, enhances the generalizability of our results to routine care settings. Generalizability is further strengthened by including patients with and without AD use at baseline, reflecting the characteristics of patients with MDD in routine clinical practice.

Our study is not without its limitations, however. The sample size was smaller than planned due to time and budgetary constraints. However, our ability to detect noninferiority based on the predefined margin for the primary outcome during and after treatment and follow-up suggests that the sample size was sufficient to assess noninferiority [73]. This noninferiority of bCBT compared with TAU was also supported by the secondary QIDS-16 outcome at the 3- and 6-month assessments, consistent with the findings for the primary outcome (PHQ-9). However, the 2 depression measures no longer yielded consistent findings at the 12-month assessment, suggesting that the longer-term treatment effects should be interpreted with some caution. Next, all participants met the diagnostic criteria for MDD (MINI) and had a PHQ-9 score above 5 at baseline. However, because treatment was delivered within routine health care settings, participants received bCBT or TAU in either primary or specialized mental health care services. Consequently, the duration and intensity of treatment varied across settings; for example, specialized mental health care generally provided more frequent and longer treatment sessions than primary care. This between-setting heterogeneity represents a potential source of confounding that is inherent to pragmatic trials. Nevertheless, leave-one-out sensitivity analyses indicated that the overall findings were robust and were not driven by any single study site. Only the noninferiority analysis of the 12-month primary outcome showed some uncertainty after exclusion of 1 site, suggesting that the longer-term findings should be interpreted with some caution. Although this heterogeneity may have influenced the observed treatment effects, it also reflects routine clinical practice and therefore enhances the external validity of the findings.

The inability to blind patients and therapists, a common feature of psychotherapeutic studies, may also have influenced our results [74]. However, the assessors of the MINI interview were blinded, and other assessments were based on self-report. It was not possible to assess the therapists' treatment fidelity, which may also have influenced the treatment outcomes. Additionally, some therapists delivered both treatments, which was necessary because the RCT was conducted within routine care. This may also have influenced treatment outcomes due to a preference for, or greater familiarity with, one treatment over the other [75]. Given that bCBT was still relatively new for most therapists, this could indeed have been an issue. Our cost-effectiveness analyses may also have been hampered by the sample size calculation, which was based on a noninferiority clinical margin rather than on economic outcomes. Thus, the sample size may have been underpowered to detect real effects in the cost differences between the 2 groups.

## Conclusions

To the best of our knowledge, this is the first large-scale RCT evaluating the clinical effectiveness, cost-effectiveness, and safety of bCBT for MDD in routine care, compared with TAU. The demonstration of clinical noninferiority positions bCBT as an additional acceptable treatment option for patients and therapists in routine mental health care. However, large-scale



XSL•FO
RenderX

implementation will also depend on more robust evidence of its cost-effectiveness. This may improve as blended care becomes more established in practice, with increasing familiarity among patients and therapists and further optimization and standardization of delivery processes. Future studies are advised to support our outcomes. Studies are also recommended to investigate in greater depth the facilitators and contextual characteristics of bCBT for successful implementation in routine care. Future studies are also needed to explore alternative modalities of "blending," such as face-to-face psychotherapy combined with videoconferencing, virtual reality, or AI-supported interaction between therapists and patients based on natural language models, alongside active and passive EMAs.

## Acknowledgments

The authors would also like to acknowledge the contributions of the patients, therapists, and general practitioners from the participating countries, as well as GAMIAN for keeping the patients' perspective at the forefront. A special tribute is reserved for Dr Jeroen Ruwaard and Professor Dr Jan Smit, who both played a pivotal role in the European Comparative Effectiveness Research on Internet-based Depression Treatment (E-COMPARED) study. Dr Jeroen Ruwaard passed away unexpectedly in 2019. Unfortunately, Professor Dr Jan Smit had to withdraw as coauthor due to personal circumstances. Finally, the authors would like to thank all other contributors to the E-COMPARED project, particularly Dr Christiaan Vis and Dr Stasja Draisma and the External Advisory Board for their reflection upon the scope and outcomes of the E-COMPARED project study. Machine tools were used to suggest language improvements within the manuscript.

## Funding

The European Comparative Effectiveness Research on Internet-based Depression Treatment (E-COMPARED) study was funded by the European Union's Seventh Framework (FP7), grant agreement HEALTH-F3-2013-603098. Denmark received funding from the Innovation Fund Denmark to support the Danish trial in the ENTER project, which provided data for the E-COMPARED project. The funding bodies did not influence the study design.

## Data Availability

Data are available on request from the authors.

## Authors' Contributions

HR, AK, JB, RA, and MH designed this study. HR and AK managed the project. AvS, KM, GA, DDE, TB, KC, RB, CB, RA, and RC led the national study sites as principal investigators. IT, LK, KM, TK, JBH, AC-P, AG-P, NT, and ES managed and conducted the national trials. The analyses were carried out by AH, JB, HR, AK, and MS, while HR, AK, AH, and JB wrote the manuscript. All authors reviewed, provided input, and approved the final manuscript.

## Conflicts of Interest

DDE reports having received consultancy fees from, and served on the scientific advisory boards of, several companies such as Sanofi, Minddistrict, Lantern, Schoen Kliniken, and German health insurance companies (Techniker Krankenkasse, BARMER). He is a shareholder of the Institute for health training online GmbH (GET.ON/HelloBetter), a provider of digital therapeutics for mental disorders in routine mental health care (unrelated to the present work). DDE report having received consultancy fees and fees for lectures/workshops from chambers of psychotherapists and training institutes for psychotherapists in the e-mental-health context. All other authors declared no conflicts of interest.

Multimedia Appendix 1
Recruitment and treatment settings by country.
[PDF File, 100 KB - i-jmr_v15i1e72866_app1.pdf ]

Multimedia Appendix 2
Reliability (Cronbach α) of the questionnaires used.
[PDF File, 139 KB - i-jmr_v15i1e72866_app2.pdf ]

Multimedia Appendix 3
Calculation of the deterioration at 3 months, 6 months, and 12 months.
[PDF File, 146 KB - i-jmr_v15i1e72866_app3.pdf ]

Multimedia Appendix 4
Baseline sociodemographic characteristics (means, SD) per country (N=835).



XSL•FO
RenderX

[PDF File, 141 KB - i-jmr_v15i1e72866_app4.pdf ]

Checklist 1
CONSORT 2010 checklist randomized controlled trial.
[PDF File, 180 KB - i-jmr_v15i1e72866_app5.pdf ]

## References


1. Rong J, Wang X, Cheng P, Li D, Zhao D. Global, regional and national burden of depressive disorders and attributable risk factors, from 1990 to 2021: results from the 2021 Global Burden of Disease study. Br J Psychiatry 2025 Oct;227(4):688-697. [doi: 10.1192/bjp.2024.266] [Medline: 39809717]
2. Hasin DS, Sarvet AL, Meyers JL, et al. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. JAMA Psychiatry 2018 Apr 1;75(4):336-346. [doi: 10.1001/jamapsychiatry.2017.4602] [Medline: 29450462]
3. Greenberg PE, Fournier AA, Sisitsky T, et al. The economic burden of adults with major depressive disorder in the United States (2010 and 2018). Pharmacoeconomics 2021 Jun;39(6):653-665. [doi: 10.1007/s40273-021-01019-4] [Medline: 33950419]
4. Thaipisuttikul P, Ittasakul P, Waleeprakhon P, Wisajun P, Jullagate S. Psychiatric comorbidities in patients with major depressive disorder. Neuropsychiatr Dis Treat 2014;10:2097-2103. [doi: 10.2147/NDT.S72026] [Medline: 25419132]
5. Cuijpers P, Quero S, Noma H, et al. Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. World Psychiatry 2021 Jun;20(2):283-293. [doi: 10.1002/wps.20860] [Medline: 34002502]
6. McHugh RK, Whitton SW, Peckham AD, Welge JA, Otto MW. Patient preference for psychological vs pharmacologic treatment of psychiatric disorders: a meta-analytic review. J Clin Psychiatry 2013 Jun;74(6):595-602. [doi: 10.4088/JCP.12r07757] [Medline: 23842011]
7. Gao Y, Burns R, Leach L, Chilver MR, Butterworth P. Examining the mental health services among people with mental disorders: a literature review. BMC Psychiatry 2024 Aug 20;24(1):568. [doi: 10.1186/s12888-024-05965-z] [Medline: 39164690]
8. Moitra M, Santomauro D, Collins PY, et al. The global gap in treatment coverage for major depressive disorder in 84 countries from 2000-2019: a systematic review and Bayesian meta-regression analysis. PLoS Med 2022 Feb;19(2):e1003901. [doi: 10.1371/journal.pmed.1003901] [Medline: 35167593]
9. Riper H, Cuijpers P. Telepsychology and eHealth. In: Norcross JC, VandenBos GR, Freedheim DK, editors. APA Handbook of Clinical Psychology: Applications and Methods: American Psychological Association; 2016:451-463. [doi: 10.1037/14861-024]
10. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019 Feb;18(1):20-28. [doi: 10.1002/wps.20610] [Medline: 30600624]
11. Harrer M, Miguel C, Tong L, et al. Metapsy consortium. Effectiveness of digital interventions for eight mental disorders: a meta-analytic synthesis. Internet Interv 2025 Sep;41:100860. [doi: 10.1016/j.invent.2025.100860] [Medline: 40703853]
12. Hedman-Lagerlöf E, Carlbring P, Svärdman F, Riper H, Cuijpers P, Andersson G. Therapist-supported Internet-based cognitive behaviour therapy yields similar effects as face-to-face therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. World Psychiatry 2023 Jun;22(2):305-314. [doi: 10.1002/wps.21088] [Medline: 37159350]
13. Hafner J, Tokgöz P, Dockweiler C. Implementation factors of digital health interventions in depression care-the perspective of health professionals. Healthcare (Basel) 2025 Oct 27;13(21):2717. [doi: 10.3390/healthcare13212717] [Medline: 41228083]
14. Mol M, van Genugten C, Dozeman E, et al. Why uptake of blended internet-based interventions for depression Is challenging: a qualitative study on therapists' perspectives. J Clin Med 2020 Jan;9(1):91. [doi: 10.3390/jcm9010091]
15. Berger T. The therapeutic alliance in internet interventions: a narrative review and suggestions for future research. Psychother Res 2017 Sep;27(5):511-524. [doi: 10.1080/10503307.2015.1119908] [Medline: 26732852]
16. Kooistra L, Ruwaard J, Wiersma J, van Oppen P, Riper H. Working alliance in blended versus face-to-face cognitive behavioral treatment for patients with depression in specialized mental health care. J Clin Med 2020 Jan 27;9(2):347. [doi: 10.3390/jcm9020347] [Medline: 32012722]
17. Ferrao Nunes-Zlotkowski K, Shepherd HL, Beatty L, Butow P, Shaw JM. Blended psychological therapy for the treatment of psychological disorders in adult patients: systematic review and meta-analysis. Interact J Med Res 2024 Oct 29;13:e49660. [doi: 10.2196/49660] [Medline: 39470720]
18. Erbe D, Eichert HC, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306. [doi: 10.2196/jmir.6588] [Medline: 28916506]




XSL•FO
RenderX

19. Sextl-Plötz T, Titzler I, Schmidt-Hantke J, et al. Patients' perspectives on implementing fixed and flexible variants of blended therapy in routine care: a qualitative study of the project PSYCHOnlineTHERAPY. Digit Health 2025;11:20552076251377943. [doi: 10.1177/20552076251377943] [Medline: 40979695]
20. Miller RK, O'Moore K, Kikas K, et al. Needs and expectations for the myNewWay blended digital and face-to-face psychotherapy model of care for depression and anxiety (part 2): participatory design study including mental health professionals. JMIR Hum Factors 2025 Dec 4;12:e68789. [doi: 10.2196/68789] [Medline: 41343853]
21. Titzler I, Saruhanjan K, Berking M, Riper H, Ebert DD. Barriers and facilitators for the implementation of blended psychotherapy for depression: a qualitative pilot study of therapists' perspective. Internet Interv 2018 Jun;12:150-164. [doi: 10.1016/j.invent.2018.01.002] [Medline: 30135779]
22. Kalde J, Atik E, Stricker JH, et al. Enhancing the efficacy of CBT for patients with unipolar depression by integrating digital interventions into treatment: a pilot randomized controlled trial. Internet Interv Preprint posted online on 2024. [doi: 10.31234/osf.io/zsfkc]
23. Kooistra LC, Wiersma JE, Ruwaard J, et al. Cost and effectiveness of blended versus standard cognitive behavioral therapy for outpatients with depression in routine specialized mental health care: pilot randomized controlled Trial. J Med Internet Res 2019 Oct 29;21(10):e14261. [doi: 10.2196/14261] [Medline: 31663855]
24. Baumeister H, Bauereiss N, Zarski AC, et al. Clinical and cost-effectiveness of PSYCHOnlineTHERAPY: study protocol of a multicenter blended outpatient psychotherapy cluster randomized controlled trial for patients with depressive and anxiety disorders. Front Psychiatry 2021;12(8):660534. [doi: 10.3389/fpsyt.2021.660534] [Medline: 34054617]
25. Massoudi B, Blanker MH, van Valen E, Wouters H, Bockting CLH, Burger H. Blended care vs. usual care in the treatment of depressive symptoms and disorders in general practice [BLENDING]: study protocol of a non-inferiority randomized trial. BMC Psychiatry 2017 Jun 13;17(1):218. [doi: 10.1186/s12888-017-1376-1] [Medline: 28610561]
26. Luyten P, Speybrouck D, Martin P, et al. Clinical and cost-effectiveness of BLENDED cognitive behavioral therapy or psychodynamic therapy versus face-to-face psychotherapy for depression (BLENDED Study): protocol for a pragmatic, multicenter, assessor-blinded randomized controlled noninferiority trial. JMIR Res Protoc 2026 Jan 14;15:e80511. [doi: 10.2196/80511] [Medline: 41533960]
27. Kleiboer A, Smit J, Bosmans J, et al. European COMPARative effectiveness research on blended depression treatment versus treatment-as-usual (E-COMPARED): study protocol for a randomized controlled, non-inferiority trial in eight European countries. Trials 2016 Aug 3;17(1):387. [doi: 10.1186/s13063-016-1511-1] [Medline: 27488181]
28. Blackwelder WC. Current issues in clinical equivalence trials. J Dent Res 2004;83 Spec No C:C113-C115. [doi: 10.1177/154405910408301s23] [Medline: 15286135]
29. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJW, CONSORT Group. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. JAMA 2006 Mar 8;295(10):1152-1160. [doi: 10.1001/jama.295.10.1152] [Medline: 16522836]
30. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
31. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20(22-33):22-33. [Medline: 9881538]
32. Cuijpers P, Turner EH, Koole SL, van Dijke A, Smit F. What is the threshold for a clinically relevant effect? The case of major depressive disorders. Depress Anxiety 2014 May;31(5):374-378. [doi: 10.1002/da.22249] [Medline: 24677535]
33. European Parliament, Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Off J Eur Union 2016;L119:1-88 [FREE Full text]
34. National Institute for Health and Care Excellence. Depression in Adults: Treatment and Management NICE Guideline NG222: National Institute for Health and Care Excellence; 2022. URL: https://www.nice.org.uk/guidance/ng222 [accessed 2026-08-12]
35. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J Gen Intern Med 2007 Nov;22(11):1596-1602. [doi: 10.1007/s11606-007-0333-y] [Medline: 17874169]
36. Wittkampf KA, Naeije L, Schene AH, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. Gen Hosp Psychiatry 2007;29(5):388-395. [doi: 10.1016/j.genhosppsych.2007.06.004] [Medline: 17888804]
37. Erbe D, Eichert HC, Rietz C, Ebert DD. Interformat reliability of the patient health questionnaire: validation of the computerized version of the PHQ-9. Internet Interv 2016 Sep;5:1-4. [doi: 10.1016/j.invent.2016.06.006] [Medline: 30135800]
38. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [doi: 10.1037/0022-006X.59.1.12]




XSL•FO
RenderX

39. Rush AJ, Trivedi MH, Ibrahim HM, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol Psychiatry 2003 Sep 1;54(5):573-583. [doi: 10.1016/s0006-3223(02)01866-8] [Medline: 12946886]
40. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry 2000 Jun;31(2):73-86. [doi: 10.1016/s0005-7916(00)00012-4] [Medline: 11132119]
41. Nguyen TD, Attkisson CC, Stegner BL. Assessment of patient satisfaction: development and refinement of a service evaluation questionnaire. Eval Program Plann 1983;6(3-4):299-313. [doi: 10.1016/0149-7189(83)90010-1] [Medline: 10267258]
42. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry: Taylor & Francis; 1996:189-194. [doi: 10.1201/9781498710411-35]
43. Hatcher RL, Gillaspy JA Jr. Development and validation of a revised short version of the working alliance inventory. Psychother Res 2006 Jan;16(1):12-25. [doi: 10.1080/10503300500352500]
44. Hatcher RL, Lindqvist K, Falkenström F. Psychometric evaluation of the working alliance inventory-therapist version: current and new short forms. Psychother Res 2020 Jul;30(6):706-717. [doi: 10.1080/10503307.2019.1677964] [Medline: 31621525]
45. Herrero R, Vara MD, Miragall M, et al. Working Alliance Inventory for Online Interventions-Short Form (WAI-TECH-SF): the role of the therapeutic alliance between patient and online program in therapeutic outcomes. Int J Environ Res Public Health 2020 Aug 25;17(17):6169. [doi: 10.3390/ijerph17176169] [Medline: 32854381]
46. Hakkaart-van Roijen L, Straten A, Donker M, Tiemens B. Manual for the Trimbos/iMTA Questionnaire for Costs Associated with Psychiatric Illness (TiC-P): Erasmus MC; 2002.
47. Hakkaart-van Roijen L, van der Linden N, Bouwmans C, Kanters T, Tan SS. Costing Manual: Methodology of Costing Research and Reference Prices for Economic Evaluations in Healthcare: Zorginstituut Nederland; 2016.
48. Dunn DT, Copas AJ, Brocklehurst P. Superiority and non-inferiority: two sides of the same coin? Trials 2018 Sep 17;19(1):499. [doi: 10.1186/s13063-018-2885-z] [Medline: 30223881]
49. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res 2007 Jun;16(3):219-242. [doi: 10.1177/0962280206074463] [Medline: 17621469]
50. Hox JJ, Moerbeek M, Schoot R. Multilevel Analysis: Techniques and Applications, 3rd edition: Routledge; 2017. [doi: 10.4324/9781315650982]
51. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol 2017 Dec 6;17(1):162. [doi: 10.1186/s12874-017-0442-1] [Medline: 29207961]
52. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Lawrence Erlbaum Associates; 1988. [doi: 10.4324/9780203771587]
53. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis: John Wiley & Sons; 2009. [doi: 10.1002/9780470743386]
54. Magnusson M, Andersen MR, Jonasson J, Vehtari A. Leave-one-out cross-validation for Bayesian model comparison. Proc Mach Learn Res 2020;108:341-351 [FREE Full text]
55. Zwaap J, Knies S, van der Meijden C, Staal P, van der Heiden L. Cost-Effectiveness in Practice (Kosteneffectiviteit in de Praktijk) [in Dutch; Archived Version]: Dutch Healthcare Institute (Zorginstituut Nederland); 2015. URL: https://web.archive.org/web/20221104172820/https://www.zorginstituutnederland.nl/binaries/zinl/documenten/rapport/2015/06/26/kosteneffectiviteit-in-de-praktijk/Kosteneffectiviteit+in+de+praktijk.pdf
56. NICE technology appraisal and highly specialised technologies guidance: the manual NICE process and methods guide PMG36. National Institute for Health and Care Excellence. 2022. URL: https://www.nice.org.uk/process/pmg36? [accessed 2026-09-11]
57. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736. [doi: 10.1007/s11136-011-9903-x]
58. M Versteegh M, M Vermeulen K, M A A Evers S, de Wit GA, Prenger R, A Stolk E. Dutch tariff for the five-level version of EQ-5D. Value Health 2016 Jun;19(4):343-352. [doi: 10.1016/j.jval.2016.01.003] [Medline: 27325326]
59. Husereau D, Drummond M, Augustovski F, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. Value Health 2022 Jan;25(1):3-9. [doi: 10.1016/j.jval.2021.11.1351] [Medline: 35031096]
60. Rubin A, Yu M. Within-group effect size benchmarks for cognitive–behavioral therapy in the treatment of adult depression. Soc Work Res 2017 Sep;41(3):135-144. [doi: 10.1093/swr/svx011]
61. Cuijpers P, Miguel C, Harrer M, et al. Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: a comprehensive meta-analysis including 409 trials with 52,702 patients. World Psychiatry 2023 Feb;22(1):105-115. [doi: 10.1002/wps.21069] [Medline: 36640411]
62. Zhang Z, Zhang L, Zhang G, Jin J, Zheng Z. The effect of CBT and its modifications for relapse prevention in major depressive disorder: a systematic review and meta-analysis. BMC Psychiatry 2018 Feb 23;18(1):50. [doi: 10.1186/s12888-018-1610-5] [Medline: 29475431]




XSL•FO
RenderX

63. Cuijpers P, Miguel C, Harrer M, Ciharova M, Karyotaki E. Does the use of pharmacotherapy interact with the effects of psychotherapy? A meta-analytic review. Eur Psychiatry 2023 Aug 3;66(1):e63. [doi: 10.1192/j.eurpsy.2023.2437] [Medline: 37534407]
64. Cuijpers P, Karyotaki E, Ciharova M, Miguel C, Noma H, Furukawa TA. The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: a meta-analysis. Acta Psychiatr Scand 2021 Sep;144(3):288-299. [doi: 10.1111/acps.13335] [Medline: 34107050]
65. Malouin-Lachance A, Capolupo J, Laplante C, Hudon A. Does the digital therapeutic alliance exist? Integrative review. JMIR Ment Health 2025 Feb 7;12:e69294. [doi: 10.2196/69294] [Medline: 39924298]
66. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: a meta-analytic synthesis. Psychotherapy (Chic) 2018 Dec;55(4):316-340. [doi: 10.1037/pst0000172] [Medline: 29792475]
67. Doukani A, Quartagno M, Sera F, et al. Comparison of the working alliance in blended cognitive behavioral therapy and treatment as usual for depression in Europe: secondary data analysis of the E-COMPARED randomized controlled trial. J Med Internet Res 2024 May 31;26:e47515. [doi: 10.2196/47515] [Medline: 38819882]
68. Berger T, Krieger T, Sude K, Meyer B, Maercker A. Evaluating an e-mental health program ("deprexis") as adjunctive treatment tool in psychotherapy for depression: results of a pragmatic randomized controlled trial. J Affect Disord 2018 Feb;227:455-462. [doi: 10.1016/j.jad.2017.11.021] [Medline: 29154168]
69. Thase ME, Wright JH, Eells TD, et al. Improving the efficiency of psychotherapy for depression: computer-assisted versus standard CBT. Am J Psychiatry 2018 Mar 1;175(3):242-250. [doi: 10.1176/appi.ajp.2017.17010089] [Medline: 28969439]
70. Schaeuffele C, Mutak A, Behr S, et al. Increasing the effectiveness of psychotherapy in routine care through transdiagnostic online modules? Randomized controlled trial investigating blended care. J Consult Clin Psychol 2026 Jan;94(1):11-25. [doi: 10.1037/ccp0000983] [Medline: 41538200]
71. Zuidgeest MGP, Goetz I, Groenwold RHH, et al. Series: pragmatic trials and real world evidence: paper 1. Introduction. J Clin Epidemiol 2017 Aug;88:7-13. [doi: 10.1016/j.jclinepi.2016.12.023] [Medline: 28549929]
72. Cooper AA, Conklin LR. Dropout from individual psychotherapy for major depression: a meta-analysis of randomized clinical trials. Clin Psychol Rev 2015 Aug;40:57-65. [doi: 10.1016/j.cpr.2015.05.001] [Medline: 26067572]
73. Cuijpers P. Are all psychotherapies equally effective in the treatment of adult depression? The lack of statistical power of comparative outcome studies. Evid Based Ment Health 2016 May;19(2):39-42. [doi: 10.1136/eb-2016-102341] [Medline: 26984413]
74. Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. Lancet 2002 Feb 23;359(9307):696-700. [doi: 10.1016/S0140-6736(02)07816-9] [Medline: 11879884]
75. Falkenström F, Markowitz JC, Jonker H, Philips B, Holmqvist R. Can psychotherapists function as their own controls? Meta-analysis of the crossed therapist design in comparative psychotherapy trials. J Clin Psychiatry 2013 May;74(5):482-491. [doi: 10.4088/JCP.12r07848] [Medline: 23146326]


## Abbreviations

**AD:** antidepressant
**bCBT:** blended cognitive behavioral therapy
**CBT:** cognitive behavioral therapy
**CHEERS:** Consolidated Health Economic Evaluation Reporting Standards
**CONSORT:** Consolidated Standards of Reporting Trials
**CSQ-8:** Client Satisfaction Questionnaire-8
***DSM-IV*:** *Diagnostic and Statistical Manual of Mental Disorders* (Fourth Edition)
**E-COMPARED:** European Comparative Effectiveness Research on Internet-based Depression Treatment
**EE:** estimated intervention effects
**EMA:** ecological momentary assessment
**ftfCBT:** face-to-face CBT
**iCBT:** internet-based cognitive behavioral therapy
**ICER:** incremental cost-effectiveness ratio
**ITT:** intention-to-treat
**MDD:** major depressive disorder
**MINI:** Mini International Neuropsychiatric Interview
**OR:** odds ratio
**PHQ-9:** Patient Health Questionnaire-9
**QALY:** quality-adjusted life years
**QIDS-SR 16:** Quick Inventory of Depressive Symptomatology Self-Report
**RCI:** Reliable Change Index
**RCT:** randomized controlled trial
**SUS :** System Usability Scale



XSL•FO
RenderX

**TAU:** treatment as usual
**TiC-P:** Trimbos/iMTA questionnaire for Costs associated with Psychiatric illness
**WAI-SF:** Working Alliance Inventory-Short Form
**WAI-TECH-SF:** Working Alliance Inventory for Online Interventions-Short Form.
**WTP:** willingness-to-pay

*Edited by M Balcarras; submitted 20.Feb.2025; peer-reviewed by A Azizi, AB Garcia, MA Zoubi, R Tripathi; revised version received 21.Jul.2026; accepted 22.Jul.2026; published 24.Sep.2026.*

*Please cite as:*
*Riper H, Kleiboer A, Hoogendoorn A, Bosmans J, Schulte M, Mathiasen K, Hazo JB, Holtzmann J, Chevreul K, Ebert DD, Titzler I, Berking M, Funk B, van Schaik A, Kemmeren L, Cieslak R, Smoktunowicz E, Maj A, Botella C, Garcia-Palacios A, Banos R, Herrero R, Andersson G, Topooco N, Vernmark K, Berger T, Krieger T, Cerga-Pashoja A, Doukani A, Araya R, Karyotaki E, Kolovos S, Hoogendoorn M, van Breda W, van de Ven P, Rocha A, Gonçalves G, Cuijpers P*
*Clinical Effectiveness and Cost-Effectiveness of Blended Treatment for Major Depression Compared With Treatment as Usual Within Routine Care in Europe: A Noninferiority Randomized Controlled Trial*
*Interact J Med Res 2026;15:e72866*
*URL: https://www.i-jmr.org/2026/1/e72866*
*doi:10.2196/72866*

© Heleen Riper, Annet Kleiboer, Adriaan Hoogendoorn, Judith Bosmans, Mieke Schulte, Kim Mathiasen, Jean Baptiste Hazo, Jerome Holtzmann, Karine Chevreul, David Daniel Ebert, Ingrid Titzler, Mathias Berking, Burkhardt Funk, Anneke van Schaik, Lise Kemmeren, Roman Cieslak, Ewelina Smoktunowicz, Anna Maj, Cristina Botella, Azucena Garcia-Palacios, Rosa Banos, Rocío Herrero, Gerhard Andersson, Naira Topooco, Kristofer Vernmark, Thomas Berger, Tobias Krieger, Arlinda Cerga-Pashoja, Asmae Doukani, Ricardo Araya, Eirini Karyotaki, Spyros Kolovos, Mark Hoogendoorn, Ward van Breda, Pepijn van de Ven, Artur Rocha, Gonçalo Gonçalves, Pim Cuijpers. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 24.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Burden of Insomnia Among Individuals With Major Depressive Disorder: Cross-Sectional Observational Study

Josh Hamilton[1], MD; Emilie Pain[2], MSc; Jesper Riise[3], MSc, PhD; Zhiheng Zhang[4], MSc, MD

[1]The Hamilton Group Behavioral Health LLC, Las Vegas, NV, United States

[2]Carenity, 1 Rue de Stockholm, Paris, France

[3]Janssen Cilag A/S, Birkerød, Denmark

[4]Janssen Scientific Affairs, Titusville, NJ, United States

**Corresponding Author:**
Emilie Pain, MSc
Carenity, 1 Rue de Stockholm, Paris, France

## Abstract

**Background:** Insomnia is a common and persistent symptom in major depressive disorder (MDD), associated with poorer outcomes and reduced quality of life (QoL). Despite its prevalence, recent patient-reported data on the burden of insomnia symptoms in MDD remains limited.

**Objective:** This study aimed to assess the burden of insomnia among individuals with MDD experiencing insomnia symptoms (MDDIS) and its impact on patient-perceived disease burden and QoL.

**Methods:** A noninterventional, cross-sectional online study was conducted from October 2022 to February 2023 in adults with self-reported MDD and insomnia symptoms, residing in the United States or 1 of 5 European countries (EU5: France, Germany, Italy, Spain, and the United Kingdom). Participants completed validated questionnaires (Patient Health Questionnaire-9 [PHQ-9] for depression severity and Insomnia Severity Index [ISI] for insomnia severity) and provided data on demographics, clinical history, treatment experience, and perceptions. Descriptive statistics were used to summarize results, and subgroup analyses were conducted based on depression status and severity.

**Results:** Among 1250 participants (United States: 500; EU5: 750), most were cisgender women (United States: 363/500, 72%; EU5: 637/750, 85%), aged ~48 years, and on average 13.4 years in the United States and 11.2 years in EU5 had passed since being diagnosed with MDD. Around 60% (United States: 312/500; EU5: 439/750) reported having moderate to severe insomnia symptoms. Most had used antidepressants (United States: 460/500, 92%; EU5: 646/750, 86%) at least 40% (United States: 198/500; EU5: 343/750) in both regions reported using sleep-related medications (over-the-counter sleep aids, dual-purpose medication, or sleep-specific medication). However, insomnia symptoms persisted, with around half of participants reporting 8 to 15 insomnia nights over 2 weeks. Insomnia symptom severity differed significantly according to depression severity and current depressive episode status. Nearly all participants reported negative consequences for daily functioning, including irritability, cognitive impairment, and reduced productivity, and approximately two-thirds identified sleep disturbances as having a major impact on their daily lives. Satisfaction with current treatments, particularly their effectiveness on sleep, was low, and 65% - 71% (488/750-354/500) of participants expressed a desire for alternative therapies.

**Conclusions:** Insomnia symptoms are highly prevalent and burdensome among patients with MDD, severely affecting daily life and perceived treatment success. Insomnia remains inadequately controlled despite high medication use. These findings emphasize the need for integrated treatment strategies targeting both mood and sleep symptoms. Given its role in depression relapse, suicidality, and treatment resistance, addressing insomnia symptoms in MDD should be a clinical priority.

***(Interact J Med Res 2026;15:e93231)*** doi:10.2196/93231

**KEYWORDS**

cross-sectional study; online study; health-related quality of life; depressive disorder; insomnia

## Introduction

### Background

Major depressive disorder (MDD) or major depression is the most commonly diagnosed form of depression [1] and was ranked by the World Health Organization as the third leading cause of disease burden in 2018, with projections placing it first by 2030 [2]. MDD is defined by a persistently low mood, loss of interest or pleasure (anhedonia), recurrent suicidal thoughts, and a range of physical and cognitive symptoms—including sleep disturbance, changes in appetite and weight, fatigue, psychomotor agitation or retardation, feelings of worthlessness,



XSL•FO
RenderX

and difficulty concentrating [3-5]. According to the *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition; *DSM-5*), diagnosis requires at least 5 of these symptoms lasting a minimum of 2 weeks [6].

The global incidence of MDD has steadily increased over the past 3 decades [7], with a sharp 27.6% rise in 2020 linked to the COVID-19 pandemic [8]. The most recent Global Burden of Disease (GBD) study estimated that 229 million people worldwide were affected by MDD in 2021, with the highest prevalence rate in the United States (4527 cases per 100,000 population; 15 million cases) and Europe (3771 cases per 100,000 population; 32 million cases) [7,9]. Globally, MDD is more prevalent in women, with a female-to-male ratio of approximately 1.6, as reported in the most recent GBD study, rising to 1.9 in the United States and Europe [7].

MDD severely impacts quality of life (QoL) not only due to its core symptoms but also because of functional limitations, social disruption, and comorbidities [3,10]. The severity of depressive symptoms directly correlates with greater QoL impairment [11-13]. Among these symptoms, sleep disturbances, especially insomnia, are among the most common, affecting around 60% of people with depression [14,15]. Insomnia symptoms are frequently reported as the main complaint [16], further deteriorating QoL by worsening energy levels and mood, and overall functioning [15,17]. Sleep disturbances in MDD are also linked to more severe symptoms—persistent depression, higher relapse risk, poorer response to traditional depression treatment, and increased risk of suicidality [18-26]. Insomnia symptoms often persist despite antidepressant treatment [27] and after remission of depressive symptoms, making individuals with MDD experiencing insomnia symptoms (MDDIS) particularly difficult to treat [28,29].

Though the literature suggests MDD and sleep disturbances are associated [16,20,30], the persistence and clinical consequences of insomnia symptoms highlight a critical need for interventions to address sleep disorders in MDDIS to improve patients' outcomes. Despite extensive research [31,32], recent data capturing patient-reported outcomes on insomnia symptoms in MDD remain limited, particularly regarding how patients perceive its impact on their lives and how this knowledge can guide the development of more effective care strategies.

### Objective

To address the absence of data on patients' perspectives regarding the impact of insomnia symptoms on patients' lives and treatment needs, we collected current patient-reported data on insomnia symptoms among individuals with MDDIS in the United States and Europe. The primary objective of this descriptive study was to assess the burden of insomnia symptoms and patient-reported QoL in this population. We also sought to understand how patients perceive the relationship between their sleep disturbances and depression, how they perceive their current MDDIS treatments, and if they discuss their insomnia symptoms with a clinician or another member of their care team.

In this paper, we present findings from this observational study that provides insights into the lived experience of MDDIS, focusing on how insomnia symptoms shape their daily functioning and well-being.

## Methods

### Study Design

This was a noninterventional, observational, and cross-sectional web-based study targeting patients who self-reported a diagnosis of depression and symptoms of insomnia.

### Study Population

Eligible participants were adults (≥18 years) residing in the United States, or in 1 of 5 European countries (EU5; France, Germany, Italy, Spain, or the United Kingdom), who self-reported a health care professional diagnosis of MDD, self-reported insomnia symptoms, and who provided informed consent. Participants who did not meet the inclusion criteria or who did not fully complete the questionnaire were excluded from the analysis. No medical records, prescriptions, or physician confirmation were requested as part of the study.

The target sample size was 1250 participants: 500 from the United States and 750 from EU5.

Participants were recruited through Carenity (ELSE CARE), an online community launched in 2011, where both patients and caregivers share experiences, access health-related information, and participate in online research studies. Carenity members were invited to participate by email. Recruitment also involved online campaigns (Facebook; Meta Platforms Inc. and Google; Alphabet Inc.). No incentives were provided for participation in the study.

Data were collected between October 2022 and February 2023.

### Ethical Considerations

The study protocol was approved by the Western Institutional Review Board–Copernicus Group (WCG IRB; approval number 20223963). All participants provided electronic informed consent prior to completing the study.

### Data Collection and Measures

#### *Overview*

Participants completed a self-administered online questionnaire. The study included 39 closed-ended questions covering participants' demographic characteristics, clinical profiles, and the impact of MDDIS on daily life. All questions were mandatory, and the study required approximately 15 minutes to complete. Study-specific questions were translated by a certified translation agency in the native language of each targeted country. The translated versions were proofread by native speakers for consistency. The Patient Health Questionnaire (PHQ-9) and the Insomnia Severity Index (ISI) questionnaires were administered using the official translated versions provided by the respective copyright holders.

#### *Sociodemographic Variables*

Participants reported their sex, year of birth, and country of residence. Sex was categorized into 5 groups: cisgender male,



XSL•FO
RenderX

cisgender female, transgender male, transgender female, and nonbinary.

### *Clinical Variables*

Participants self-reported a health care professional–confirmed diagnosis of MDD and provided information on the date of diagnosis, the type of health care professional having diagnosed MDD, their comorbidities, MDD characteristics, current and past depression treatments, insomnia symptoms characteristics, and sleep treatment history.

The PHQ-9, a 9-item patient-reported instrument for assessing depression severity, was administered using validated language versions that have demonstrated good psychometric properties [33-37]. The total score ranges from 0 to 27, with higher scores indicating greater depression severity. Severity is categorized as: minimal (0-4), mild (5-9), moderate (10-14), moderately severe (15-19), and severe (20-27).

### *Outcome Variables*

In the questionnaire, participants were informed that "Insomnia, or trouble sleeping, can include falling asleep, staying asleep or waking up early, and being unable to go back to sleep."

To evaluate the burden of insomnia symptoms in MDDIS, participants were asked how many nights they had experienced insomnia symptoms or used treatments for them in the 2 weeks prior to study inclusion (for those with a current depressive episode) or prior to their most recent episode. These are referred to as "insomnia nights" in the patient questionnaire and the paper. Participants also reported the duration of their insomnia symptoms or treatment use. To measure nocturnal and daytime symptoms of insomnia and assess its severity, the ISI, a validated 7-item scale, was also administered [38]. The total score ranges from 0 to 28, with higher scores indicating greater insomnia symptom severity. Insomnia symptoms severity is categorized as: no insomnia symptoms (0-7), subthreshold (8-14), moderate (15-21), and severe (22-28).

To evaluate the impact of MDDIS on patients' QoL, participants were asked to report, through 2 closed-ended self-designed questions, the MDDIS symptoms that had the greatest impact on their global daily life, and the consequences of sleep disruptions on their daily life during the most recent episode.

Participants were also asked to report their perceptions of the relationship between their MDD and insomnia symptoms, their satisfaction with treatments for MDD and insomnia symptoms, and their unmet needs regarding these treatments.

### Statistical Analysis

The study population included all participants who met the study selection criteria and completed the questionnaire. Descriptive statistics were used to characterize the study population and summarize study responses. Quantitative data are reported using mean and its 95% CI and qualitative data are reported as frequencies and percentages. Results are presented separately by region of residence (United States or EU5). Descriptive subgroup analyses were conducted according to current depressive episode status, depression severity, and other relevant clinical characteristics. Differences in the distribution of categorical variables between subgroups were assessed using Pearson chi-square tests or Fisher exact test. All statistical tests were 2-sided, and a $P$ value <.05 was considered statistically significant.

All data management and statistical analysis were performed using R software (version 4.1.2; R Core Team).

## Results

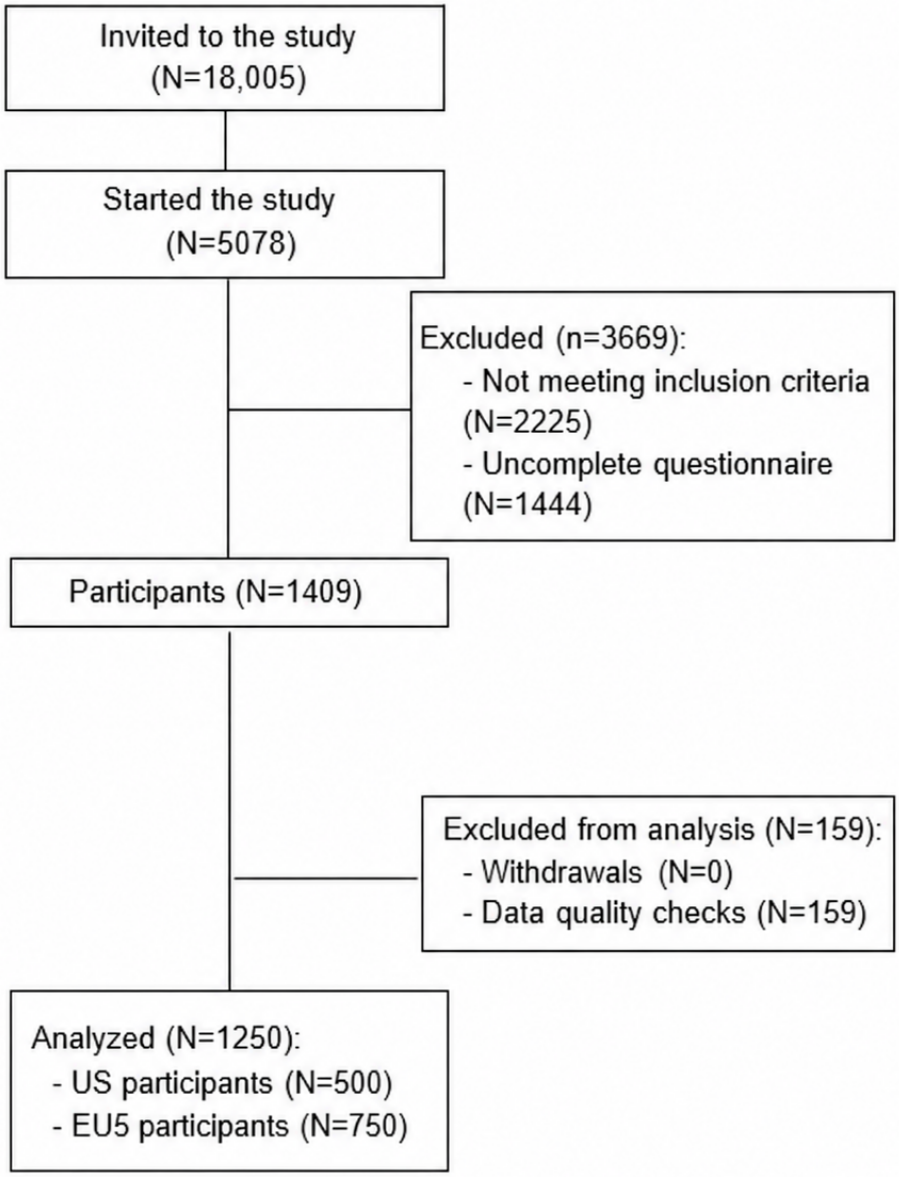
### Study Population Characteristics

Recruitment targets were met in both regions (United States and EU5), resulting in a total of 1250 participants: 500 in the United States and 750 in EU5. Details of participant disposition are shown in Figure 1. Demographic and clinical characteristics of the study population are detailed in Table 1.



XSL•FO
RenderX

**Figure 1.** Flowchart of the study population. US: United States; EU5: France, Germany, Italy, Spain, and the United Kingdom.




In both the United States and EU5, the majority of participants were cisgender women (United States: 363/500, 72%, and EU5: 637/750, 85%). A total of 12% (62/500) of US participants identified as transgender or nonbinary, compared to 3% (21/750) in the EU5 sample. Mean age at inclusion was similar across regions (United States: 48.1, 95% CI 46.5 - 49.7 years; EU5: 47.5, 95% CI 46.5 - 48.6 years) and most participants were diagnosed with MDD in their mid-thirties. In both regions, the majority of participants reported receiving their MDD diagnosis from a psychiatrist, with 79% (127/161) in Italy, 72% (111/154) in Germany, 68% (95/140) in Spain, 59% (98/167) in France, and 28% (36/128) in the United Kingdom (Table S1 in Multimedia Appendix 1).



XSL•FO
RenderX

**Table .** Demographic and health characteristics of the study population.

| Participants characteristics | US participants (n=500), n (%) | EU5[a] participants (n=750), n (%) |
|---|---|---|
| Demographic variables | | |
| Sex | | |
| Cisgender male | 68 (14) | 68 (9) |
| Cisgender female | 363 (72) | 637 (85) |
| Transgender male | 24 (5) | 6 (1) |
| Transgender female | 3 (1) | 0 (0) |
| Nonbinary | 35 (7) | 15 (2) |
| Prefer not to answer | 7 (1) | 24 (3) |
| Age (years), mean (95% CI) | 48.1 (46.5-49.7) | 47.5 (46.5-48.6) |
| Age groups (years), n (%) | | |
| 18 - 30 | 127 (25) | 123 (16) |
| 31 - 60 | 217 (43) | 499 (67) |
| >60 | 156 (31) | 128 (17) |
| Disease variables | | |
| Time since MDD[b] diagnosis (years), mean (95% CI) | 13.4 (12.2-14.5) | 11.2 (10.4-12.0) |
| Time since MDD diagnosis (years), n (%) | | |
| 0 - 5 | 201 (40) | 338 (45) |
| 6 - 20 | 153 (31) | 244 (33) |
| >20 year | 146 (29) | 168 (22) |
| Age at MDD diagnosis (years), mean (95% CI) | 34.7 (33.2 - 36.2) | 36.3 (35.3 - 37.4) |
| Age at MDD diagnosis (years), n (%) | | |
| ≤18 | 98 (20) | 90 (12) |
| 19 - 30 | 128 (26) | 207 (28) |
| 31 - 60 | 229 (46) | 415 (55) |
| >60 | 45 (9) | 38 (5) |
| HCPs[c] who diagnosed MDD, n (%) | | |
| Psychiatrist | 292 (58) | 467 (62) |
| General practitioner | 183 (37) | 237 (32) |
| Others[d] | 25 (5) | 46 (6) |
| Mental health comorbidities, n (%) | | |
| Generalized anxiety disorder | 306 (61) | 372 (50) |
| Chronic anxiety (stress) | 242 (48) | 376 (50) |
| Posttraumatic stress disorder | 202 (40) | 181 (25) |
| Social anxiety disorder | 191 (38) | 233 (31) |
| Panic disorder (or panic attacks) | 179 (36) | 256 (34) |
| Attention deficit disorder with or without hyperactivity | 98 (19) | 40 (6) |
| Eating disorder | 75 (15) | 164 (22) |
| Any personality disorder | 56 (11) | 109 (14) |
| Obsessive compulsive disorder | 67 (13) | 61 (9) |



XSL•FO
RenderX

| Participants characteristics | US participants (n=500), n (%) | EU5[a] participants (n=750), n (%) |
|---|---|---|
| Others | 55 (11) | 109 (14) |
| None | 41 (8) | 57 (9) |
| Current depressive episode, n (%) | | |
| Yes | 259 (52) | 448 (60) |
| Not sure but depression symptoms | 170 (34) | 207 (28) |
| Time elapsed since the current depressive episode[e] (years), n (%) | | |
| ≤1 | 240 (56) | 374 (57) |
| 1-5 | 97 (23) | 155 (24) |
| >5 | 92 (21) | 122 (19) |
| Unknown | 0 (0) | 4 (1) |
| No | 71 (14) | 95 (13) |
| Time elapsed since the most recent depressive episode (years), n (%) | | |
| ≤1 | 56 (79) | 52 (55) |
| 1-5 | 10 (14) | 22 (23) |
| >5 | 5 (7) | 21 (22) |
| Duration of the most recent depressive episode, n (%) | | |
| ≤6 months | 35 (49) | 39 (41) |
| 6-12 months | 19 (27) | 31 (33) |
| >1 year | 17 (24) | 25 (27) |

[a]EU5: France, Germany, Italy, Spain, and the United Kingdom.

[b]MDD: major depressive disorder.

[c]HCP: health care practitioner.

[d] "Others" includes psychologist, therapist, mental health service, mental health counselor, nurse practitioner, neurologist, counseling center, not specified.

[e]Time elapsed since the current episode applies to participants with current MDD and those who reported experiencing symptoms of depression.

Comorbid mental health conditions were prevalent. The majority of participants from both regions reported at least one additional condition, most commonly generalized anxiety disorder, chronic anxiety or stress, posttraumatic stress disorder, social anxiety disorder, and panic disorder. At the time of the study, 86% (429/500) of United States and 87% (655/750) of EU5 participants were experiencing depressive symptoms or were in an active depressive episode, which had started less than a year earlier for over half of the participants. Among those not in an active episode, most had experienced one within the past year.

### *Depression Severity*

As measured by the PHQ-9, 34% (170/500) of US participants were classified as having severe depression, 25% (124/500) as moderately severe, and 25% (123/500) as moderate depression (Figure 2). In the EU5 region, 41% (306/750) of participants had severe depression, 27% (203/750) moderately severe, and 19% (146/750) moderate depression. Descriptively, the proportion of participants with severe depression was highest in Spain, the United Kingdom, and Germany, and lower in France and Italy (Table S1 in Multimedia Appendix 1). Across both regions, the distribution of depression severity differed according to age group (both $P<.001$) and the number of insomnia nights reported during the previous 2 weeks (United States $P<.001$; EU5 $P=.04$), with a higher proportion of younger participants (18-30 years) and those who had experienced more insomnia nights reporting severe depression (Tables 2 and 3). Among US participants, a greater proportion of transgender or nonbinary participants had severe depression ($P=.02$).



XSL•FO
RenderX

**Figure 2.** Severity of depression in US (n=500) and EU5 (n=750) participants as measured by the Patient Health Questionnaire-9. EU5: France, Germany, Italy, Spain, and the United Kingdom; PHQ-9: Patient Health Questionnaire-9.

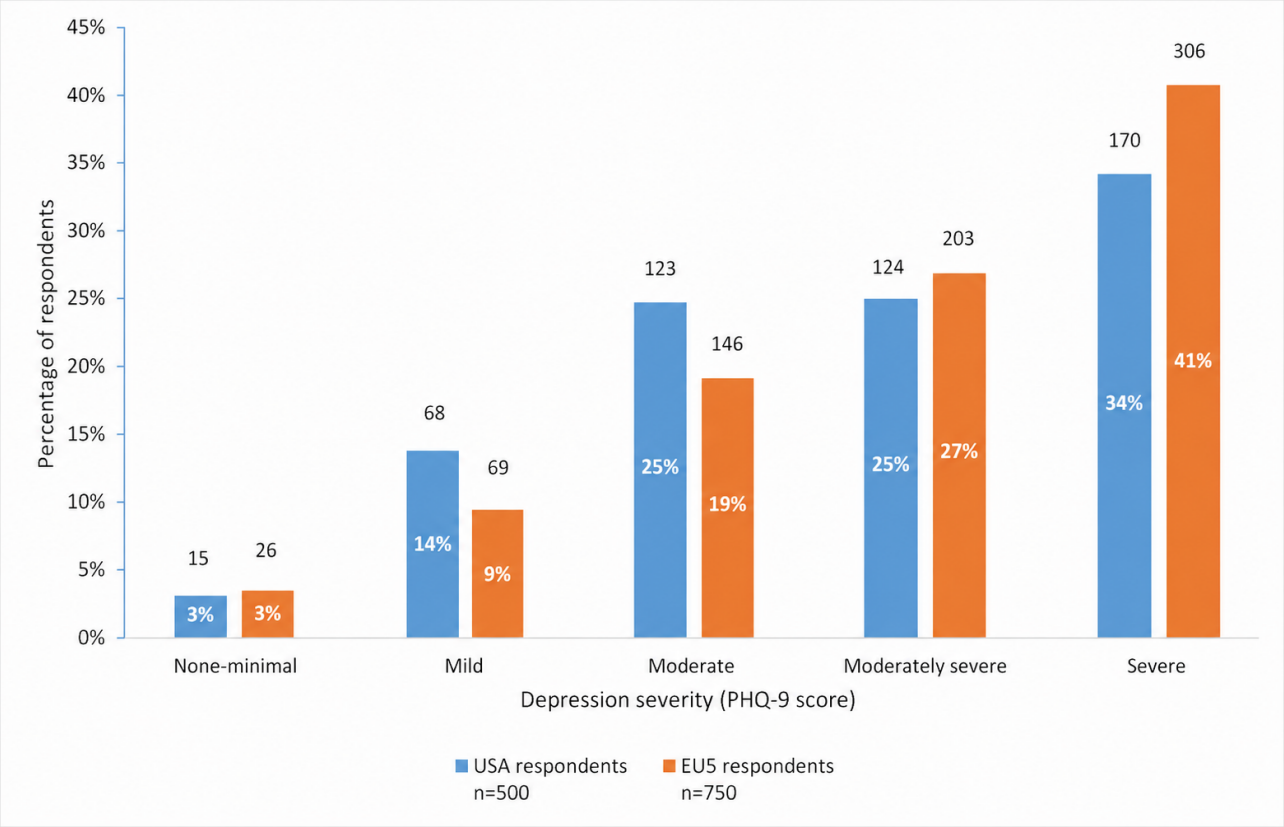



**Table .** Depression severity by sex, age, and insomnia symptoms frequency in US participants (n=500) as measured by the PHQ-9[a].

| | None-minimal (n=15) | Mild (n=68) | Moderate (n=123) | Moderately severe (n=124) | Severe (n=170) | *P* values |
|---|---|---|---|---|---|---|
| Sex, n (%) | | | | | | .02[b] |
| Cisgender female | 13 (4) | 54 (15) | 91 (25) | 92 (25) | 113 (31) | |
| Cisgender male | 2 (3) | 10 (15) | 17 (25) | 20 (29) | 19 (28) | |
| Transgender male and female and nonbinary | 0 (0) | 4 (6) | 15 (22) | 12 (17) | 38 (55) | |
| Age groups (years), n (%) | | | | | | <.001 |
| 18 - 30 | 2 (2) | 8 (6) | 23 (18) | 27 (21) | 67 (53) | |
| 31 - 60 | 8 (4) | 37 (17) | 44 (20) | 58 (27) | 70 (32) | |
| >60 | 5 (3) | 23 (15) | 56 (36) | 39 (25) | 33 (21) | |
| Insomnia nights[c], n (%) | | | | | | <.001 |
| 0 - 7 | 8 (4) | 37 (20) | 49 (26) | 32 (17) | 62 (33) | |
| 8 - 15 | 2 (1) | 6 (6) | 30 (18) | 56 (34) | 68 (41) | |

[a]PHQ-9: Patient Health Questionnaire-9.

[b]Fisher exact test.

[c]Insomnia nights are defined here as nights with insomnia symptoms or taking treatments for insomnia symptoms (for n=276 for 8 - 15 insomnia nights; for the United States, n=188 for 0 - 7 insomnia nights and n=166 for 8 - 15 insomnia nights).



XSL•FO
RenderX

**Table .** Depression severity by sex, age, and insomnia symptoms frequency in EU5[a] (N=750) participants as measured by the PHQ-9[b].

| | None-minimal (n=26) | Mild (n=69) | Moderate (n=146) | Moderately severe (n=203) | Severe (n=306) | *P* values |
|---|---|---|---|---|---|---|
| Sex, n (%) | | | | | | .45[c] |
| Cisgender female | 21 (3) | 60 (9) | 122 (19) | 175 (27) | 259 (42) | |
| Cisgender male | 5 (7) | 7 (10) | 16 (24) | 16 (24) | 24 (35) | |
| Transgender male and female and nonbinary | 0 (0) | 2 (4) | 8 (18) | 12 (27) | 23 (51) | |
| Age groups (years), n (%) | | | | | | <.001 |
| 18 - 30 | 2 (2) | 2 (2) | 16 (13) | 32 (26) | 71 (58) | |
| 31 - 60 | 17 (3) | 47 (9) | 100 (20) | 137 (27) | 198 (40) | |
| >60 | 7 (5) | 20 (16) | 30 (23) | 34 (27) | 37 (29) | |
| Insomnia nights[d], n (%) | | | | | | .04 |
| 0 - 7 | 4 (2) | 21 (10) | 38 (18) | 58 (28) | 85 (41) | |
| 8 - 15 | 6 (2) | 12 (4) | 39 (14) | 76 (28) | 143 (52) | |

[a]EU5: France, Germany, Italy, Spain, and the United Kingdom.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]Fisher exact test.

[d]Insomnia nights are defined here as nights with insomnia symptoms or taking treatments for insomnia symptoms (for EU5, n=206 for 0 - 7 insomnia nights and n=276 for 8 - 15 insomnia nights; n=166 for 8 - 15 insomnia nights).

### *Burden of Depressive Symptoms*

At least half of the participants considered each PHQ-9 depressive symptom burdensome. In both regions, the most frequently reported disturbing symptoms, reported by 93% or more of participants, were low energy (United States: 491/500 and EU5: 731/750), depressed mood (United States: 478/500 and EU5: 722/750), sleep issues (United States: 475/500 and EU5: 701/750), and anhedonia (United States: 471/500 and EU5: 714/750; Figures 3A and 3B). Approximately half of the participants in both regions were impacted by sleep disturbances nearly every day.



XSL•FO
RenderX

**Figure 3.** Burden of depressive symptoms in United States (A, n=500) and EU5 (B, n=750) participants, PHQ-9: Patient Health Questionnaire.

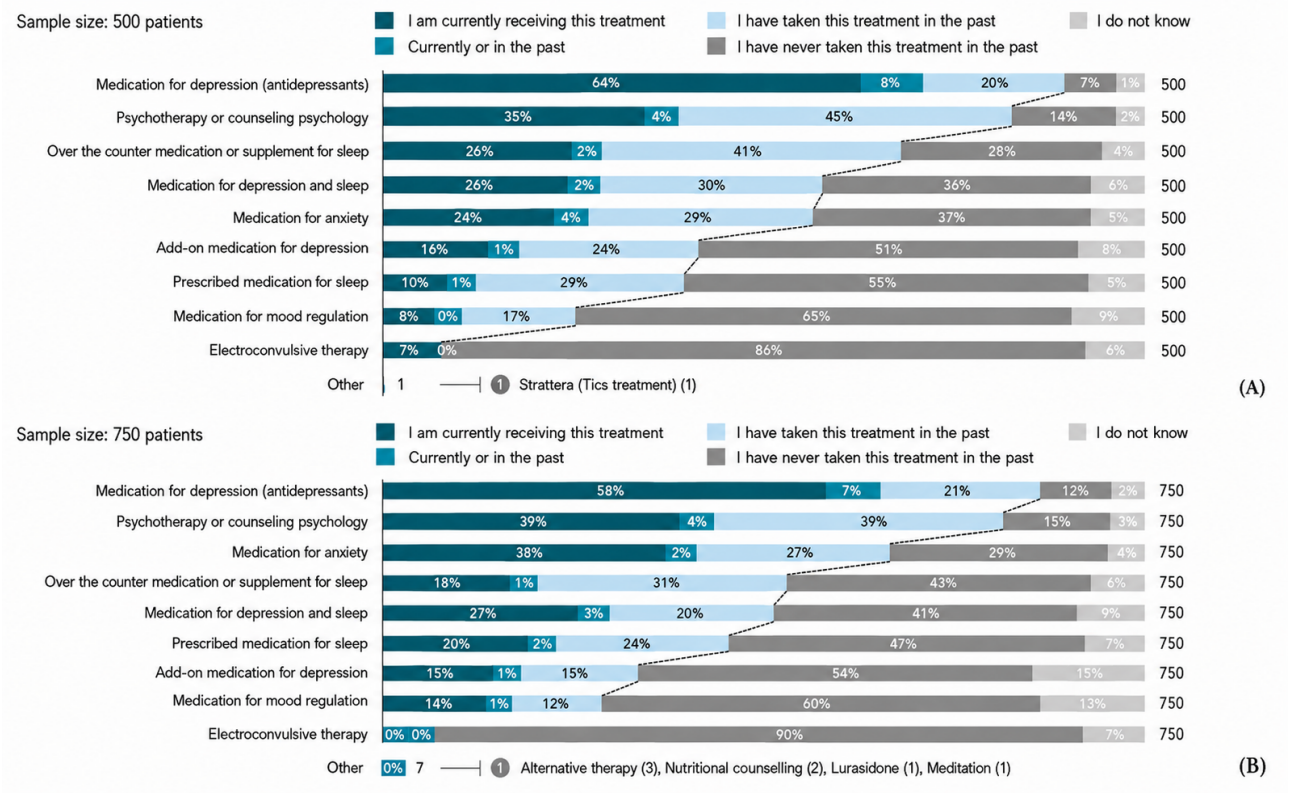
### *MDDIS Treatment*

Most participants had received antidepressant therapy (United States: 460/500, 92% and EU5: 646/750, 86%) and psychological care (United States: 421/500, 84% and EU5: 619/750, 82%) (Figure 4). Sleep-related treatments were also common. In the United States, 68% (341/500) reported using or having used over-the-counter (OTC) sleep aids, 58% (292/500) medications targeting both sleep and depression (dual-purpose medication), and 40% (198/500) prescribed sleep-specific medication. In the EU5, these figures were 51% (382/750), 50% (374/750), and 46% (343/750), respectively. At the time of the study, 28% (138/500 and 140/500) of US participants were taking OTC or dual-purpose medication and 11% (54/500) prescribed sleep-specific treatment. In the EU5, 19% (146/750) were using OTC sleep aids, 30% (222/750) used dual-purpose medication, and 22% (161/750) were receiving prescribed sleep-specific treatment.



XSL•FO
RenderX

**Figure 4.** History of treatment for US (N=500) and EU5 (N=750) participants with major depressive disorder (MDD) with insomnia symptoms. EU5: France, Germany, Italy, Spain, and the United Kingdom.




During their current or most recent depressive episode, 36% (160/438) of US participants and 30% (193/643) of EU5 participants reported changing depression treatment 3 times or more.

## Burden of Insomnia Symptoms in MDDIS and Its Impact on Patient-Reported Outcomes and QoL

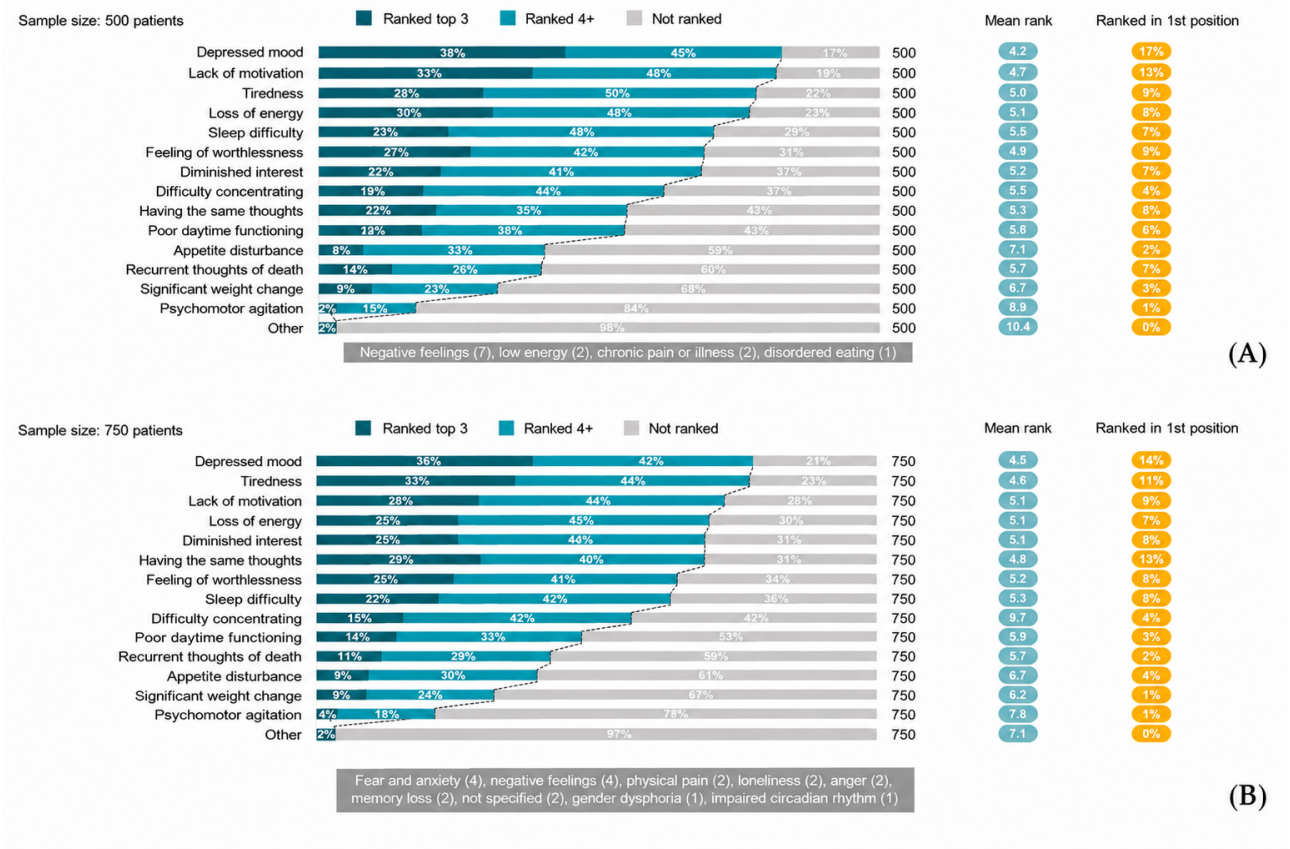
### *Insomnia Symptoms Frequency*

In the US sample, participants reported a mean of 8.1 (95% CI 7.6-8.6) insomnia nights over the 2 previous weeks. Most (354/500, 71%) reported life-impacting sleep difficulties (Figure 5) and had experienced insomnia symptoms for less than 6 months (307/500, 61%; Table 4). The distribution of insomnia-night categories (0-7 vs 8-15 nights) differed according to current depressive episode status ($P$=.05) and depression severity ($P$<.001). In particular, the proportion of US participants reporting 8-15 insomnia nights was higher among those with moderately severe (56/88, 64%) or severe depression (68/130, 52%) than among those with lower depression severity categories. The distribution of insomnia-night categories also differed according to sleep medication use ($P$=.04), with a higher proportion of participants reporting 8-15 insomnia nights among those reporting using OTC or prescribed sleeping medication (153/312, 49% vs 13/42, 31% for nonsleeping medication users).



XSL•FO
RenderX

**Figure 5.** Impact of major depressive disorder with insomnia symptoms on participants' life in the US (A, n=500) and EU5 (n=750) participants. EU5: France, Germany, Italy, Spain, and the United Kingdom.






XSL•FO
RenderX

**Table .** Number of insomnia nights in the past 2 weeks by depression status, severity, sleep medication use, and duration of insomnia symptoms in United States (n=354) and EU5[a] (n=482) participants.

| Characteristics | US participants | EU5 participants |
|---|---|---|
| Number of insomnia nights[b], mean (95% CI) | 8.1 (7.6 - 8.6) | 9.0 (8.6 - 9.4) |
| Number of insomnia nights, depressive episode status groups | | |
| Currently in a depressive episode, n | 310 | 434 |
| 0 - 7 nights, n (%) | 158 (51) | 186 (43) |
| 8 - 15 nights, n (%) | 152 (49) | 248 (57) |
| Not currently in a depressive episode, n | 44 | 48 |
| 0 - 7 nights, n (%) | 30 (68) | 20 (42) |
| 8 - 15 nights, n (%) | 14 (32) | 28 (58) |
| *P* value | .047 | .99 |
| Number of insomnia nights, depression severity groups | | |
| None-minimal, n | 10 | 10 |
| 0 - 7 nights, n (%) | 8 (80) | 4 (40) |
| 8 - 15 nights, n (%) | 2 (20) | 6 (60) |
| Mild, n | 43 | 33 |
| 0 - 7 nights, n (%) | 37 (86) | 21 (64) |
| 8 - 15 nights, n (%) | 6 (14) | 12 (36) |
| Moderate, n | 79 | 77 |
| 0 - 7 nights, n (%) | 49 (62) | 38 (49) |
| 8 - 15 nights, n (%) | 30 (38) | 39 (51) |
| Moderately severe, n | 88 | 134 |
| 0 - 7 nights, n (%) | 32 (36) | 58 (43) |
| 8 - 15 nights, n (%) | 56 (64) | 76 (57) |
| Severe, n | 130 | 228 |
| 0 - 7 nights, n (%) | 62 (48) | 85 (37) |
| 8 - 15 nights, n (%) | 68 (52) | 143 (63) |
| *P* value | <.001 | .04 |
| Number of insomnia nights, sleep medication groups | | |
| Over-the-counter or prescribed sleeping medication, n | 312 | 387 |
| 0 - 7 nights, n (%) | 159 (51) | 160 (41) |
| 8 - 15 nights, n (%) | 153 (49) | 227 (59) |
| No sleeping medication, n | 42 | 95 |
| 0 - 7 nights, n (%) | 29 (69) | 46 (48) |
| 8 - 15 nights, n (%) | 13 (31) | 49 (52) |
| *P* value | .04 | .26 |
| Duration of insomnia symptoms or treatment, N | 500 | 750 |
| <6 months, n (%) | 307 (61) | 474 (63) |
| 6 months-1 year, n (%) | 97 (19) | 115 (15) |
| >1 year, n (%) | 93 (19) | 160 (21) |
| Unknown, n (%) | 3 (1) | 1 (0) |

[a]EU5: France, Germany, Italy, Spain, and the United Kingdom.



XSL•FO
RenderX

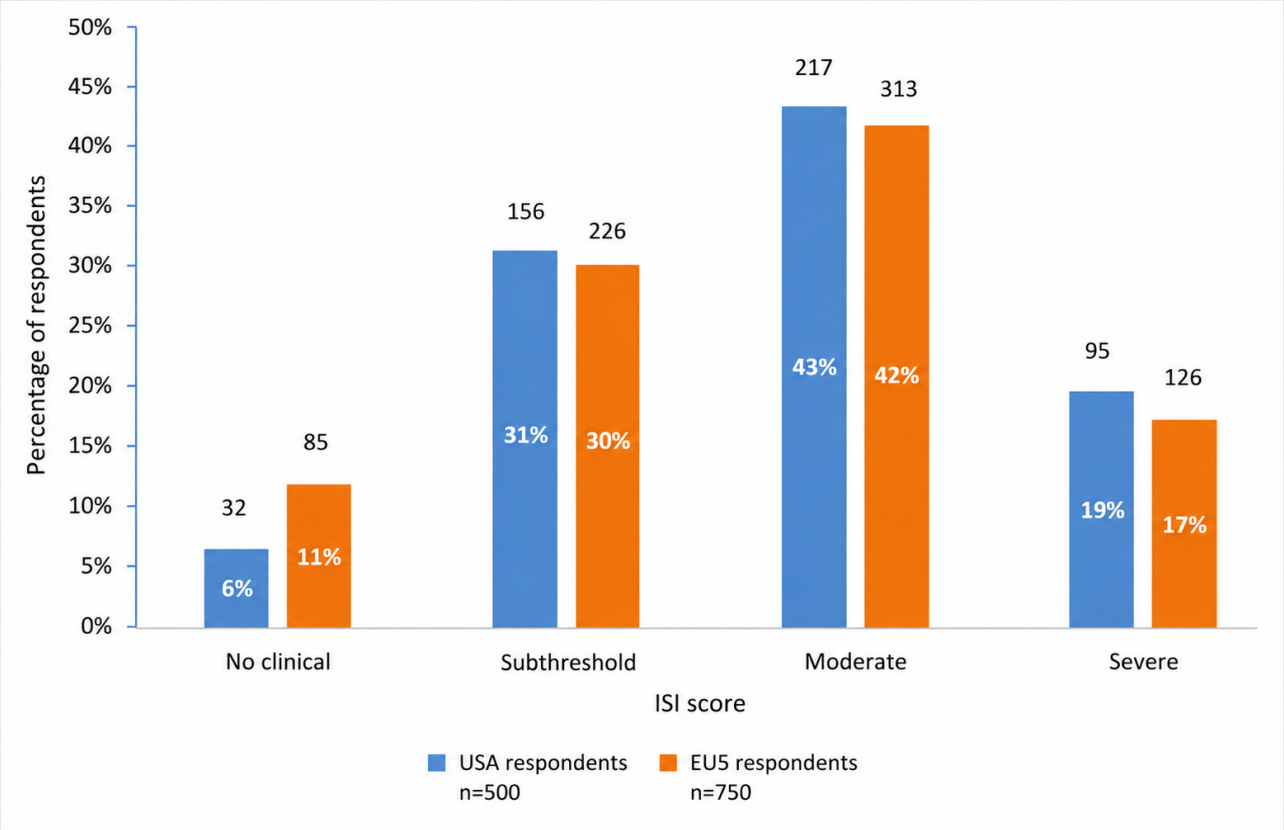
[b]Insomnia nights are defined as nights with insomnia symptoms or taking treatments for insomnia symptoms.

In the EU5 sample, an average of 9.0 (95% CI 8.6 - 9.4) insomnia nights were reported by 482/750 (64%) participants with sleep disturbances impacting their life, and 63% (474/750) had experienced insomnia symptoms for less than 6 months (Figure 5 and Table 4). The distribution of insomnia-night categories differed across depression severity categories ($P$=.04). However, no statistically significant differences were observed according to current depressive episode status or sleep medication use. The majority of EU5 participants, including those receiving OTC or prescribed sleeping medications, continued to report 8-15 insomnia nights over the previous 2 weeks.

### *Insomnia Symptoms Severity*

Based on ISI scores, 62% (312/500) of the US and 59% (439/750) of the EU5 participants had moderate to severe insomnia symptoms (Figure 6). A total of 26% (37/140) of participants in Spain, 20% (25/128) in the United Kingdom, 18% (27/154) in Germany, 16% (26/167) in France, and 7% (11/161) in Italy reported severe insomnia symptoms (Table S1 in Multimedia Appendix 1). In both regions, the distribution of insomnia severity categories differed significantly across depression severity categories and according to current depressive episode status (all $P$<.001; Tables 5 and 6). A greater proportion of participants in the more severe depression categories were classified as having moderate or severe insomnia symptoms, with 33% (57/170) of US participants and 30% (92/306) of EU5 participants with severe depression reporting severe insomnia symptoms. Additionally, 71% of both United States (185/259) and EU5 (315/448) participants currently experiencing a depressive episode were classified as having moderate or severe insomnia symptoms. In both regions, 40% (United States: 200/500 and EU5: 300/750) of participants considered insomnia symptoms interfered with their daily functioning, and more than half (279/500, 56% of United States and 437/750, 58% of EU5 participants) believed that this impairment was noticeable to others. Dissatisfaction with sleep patterns was reported by 77% (385/500) of US participants and 68% (509/750) of those in the EU5, while 34% (167/500; United States) and 30% (219/750; EU5) expressed important distress or concern about their sleep (Table 7).

**Figure 6.** Severity of insomnia symptoms in US (n=500) and EU5 (n=750) participants as measured by the Insomnia Severity Index. EU5: France, Germany, Italy, Spain, and the United Kingdom; ISI: insomnia severity index.






XSL•FO
RenderX

**Table .** Severity of insomnia symptoms by depression status and severity in US participants (n=500) as measured by the ISI[a].

| | No clinical insomnia symptoms (n=32) | Mild to moderate (n=156) | Moderate (n=217) | Severe (n=95) | *P* value |
|---|---|---|---|---|---|
| Current depression status, n (%) | | | | | <.001 |
| Not currently in a depressive episode (n=71) | 12 (17) | 33 (46) | 20 (28) | 69 (9) | |
| Unsure but experiencing depression symptoms (n=170) | 10 (6) | 59 (35) | 79 (46) | 22 (13) | |
| Currently in a depressive episode (n=259) | 10 (4) | 64 (25) | 118 (45) | 67 (26) | |
| Depression severity, n (%) | | | | | <.001[b] |
| None or minimal (n=15) | 5 (33) | 8 (53) | 2 (13) | 0 (0) | |
| Mild (n=68) | 9 (13) | 34 (50) | 21 (31) | 4 (6) | |
| Moderate (n=123) | 9 (7) | 48 (39) | 56 (45) | 10 (8) | |
| Moderately severe (n=124) | 7 (6) | 31 (25) | 62 (50) | 24 (19) | |
| Severe (n=170) | 2 (1) | 35 (20) | 76 (45) | 57 (33) | |

[a]ISI: Insomnia Severity Index.

[b]Fisher exact test.

**Table .** Severity of insomnia symptoms by depression status and severity in EU5[a] (N=750) participants as measured by the ISI[b].

| | No clinical insomnia symptoms, n=85 | Mild to moderate, n=226 | Moderate, n=313 | Severe, n=126 | *P* values |
|---|---|---|---|---|---|
| Current depression status, n (%) | | | | | <.001 |
| Not currently in a depressive episode (n=95) | 30 (32) | 36 (38) | 23 (24) | 6 (6) | |
| Unsure but experiencing depression symptoms (n=207) | 31 (15) | 81 (39) | 77 (37) | 18 (9) | |
| Currently in a depressive episode (n=448) | 24 (5) | 109 (24) | 213 (48) | 102 (23) | |
| Depression severity, n (%) | | | | | <.001 |
| None or minimal (n=26) | 18 (69) | 6 (23) | 1 (4) | 1 (4) | |
| Mild (n=69) | 21 (30) | 33 (48) | 13 (19) | 2 (3) | |
| Moderate (n=146) | 24 (16) | 64 (44) | 52 (36) | 6 (4) | |
| Moderately severe (n=203) | 14 (7) | 60 (29) | 104 (51) | 25 (12) | |
| Severe (n=306) | 8 (3) | 63 (20) | 143 (47) | 92 (30) | |

[a]EU5: France, Germany, Italy, Spain, and the United Kingdom.

[b]ISI: Insomnia Severity Index.



XSL•FO
RenderX

**Table .** Insomnia Severity Index.

| ISI[a] domains | US participants (n=500), n (%) | EU5[b] participants (n=750), n (%) |
|---|---|---|
| Difficulty initiating sleep | | |
| None | 28 (6) | 77 (10) |
| Mild | 64 (13) | 135 (18) |
| Moderate | 202 (40) | 238 (32) |
| Severe | 135 (27) | 209 (28) |
| Very severe | 71 (14) | 91 (12) |
| Difficulty staying asleep | | |
| None | 39 (8) | 83 (11) |
| Mild | 70 (14) | 130 (17) |
| Moderate | 172 (34) | 248 (33) |
| Severe | 139 (28) | 205 (27) |
| Very severe | 80 (16) | 84 (11) |
| Problems waking up too early | | |
| None | 89 (18) | 145 (19) |
| Mild | 101 (20) | 141 (19) |
| Moderate | 145 (29) | 202 (27) |
| Severe | 109 (22) | 168 (22) |
| Very severe | 56 (11) | 94 (13) |
| Satisfaction with current sleep pattern | | |
| Very satisfied | 9 (2) | 14 (2) |
| Satisfied | 19 (4) | 65 (9) |
| Moderately satisfied | 87 (17) | 162 (22) |
| Dissatisfied | 203 (41) | 290 (39) |
| Very dissatisfied | 182 (36) | 219 (29) |
| Noticeability of impairment attributed to the sleep problem | | |
| Not at all | 15 (3) | 42 (6) |
| A little | 56 (11) | 87 (12) |
| Somewhat | 150 (30) | 184 (25) |
| Much | 173 (35) | 250 (33) |
| Very much | 106 (21) | 187 (25) |
| Degree of distress or concern caused by the sleep problem | | |
| Not at all | 78 (16) | 147 20) |
| A little | 112 (22) | 180 (24) |
| Somewhat | 143 (29) | 204 (27) |
| Much | 104 (21) | 154 (21) |
| Very much | 63 (13) | 65 (9) |
| Interference with daily functioning | | |
| Not at all | 27 (5) | 88 (12) |
| A little | 100 (20) | 159 (21) |
| Somewhat | 173 (35) | 203 (27) |
| Much | 111 (22) | 206 (27) |
| Very much | 89 (18) | 94 (13) |



XSL•FO
RenderX

[a]ISI: Insomnia Severity Index.

[b]EU5: France, Germany, Italy, Spain, and the United Kingdom.

### *Consequences of Insomnia Symptoms on Daily Functioning and QoL*

Nearly all participants (489/494, 99% of United States and 692/730, 95% of EU5) reported at least one consequence of insomnia symptoms on their daily lives, with an average of 2.8 (95% CI 2.6 - 2.9) in the United States and 2.5 (95% CI 2.4 - 2.6) in the EU5 region. Approximately 70% (360/494, 73% in United States and 507/730, 69% in EU5) of participants reported more than one consequence, most commonly irritability, concentration difficulties, and reduced productivity (Table 8). The reported consequences of insomnia symptoms differed according to current depressive episode status for certain outcomes (Tables 8 and 9). In the US sample, participants currently experiencing a depressive episode were more likely to report reduced productivity (175/256, 68% vs 19/69, 27%; *P*<.001) and concentration difficulties (160/256, 62% vs 28/69, 40%; *P*=.004) than those not currently experiencing a depressive episode. No significant differences were observed for irritability, increased caffeine consumption, daytime sleep episodes, other consequences, or reporting no consequences. In the EU5 sample, significant differences were only observed for irritability (*P*=.006) and concentration difficulties (*P*=.02). Participants currently experiencing a depressive episode were more likely to report irritability (320/443, 72% vs 48/87, 55%) and concentration difficulties (272/443, 61% vs 40/87, 46%) than those not currently experiencing a depressive episode.

**Table .** Consequences of insomnia symptoms on daily life in US participants (n=494) with an ISI[a] score >0**.** Participants were asked "What are/were the consequences of your sleep difficulty symptom(s) on your daily life during your most recent depressive episode?"

| Consequences of insomnia symptoms[b] | US participants with ISI score >0 | | | | |
|---|---|---|---|---|---|
| | Not currently in a depressive episode (n=69) | Unsure but experiencing depression symptoms (n=169) | Currently in a depressive episode (n=256) | Total (n=494) | *P* values |
| Irritability, n (%) | 44 (64) | 132 (78) | 182 (71) | 358 (72) | .06 |
| Reduced productivity, n (%) | 19 (27) | 103 (61) | 175 (68) | 297 (60) | <.001 |
| Concentration difficulties, n (%) | 28 (40) | 106 (63) | 160 (62) | 294 (60) | .004 |
| Increased caffeine consumption, n (%) | 26 (38) | 91 (54) | 121 (47 | 238 (48) | .07 |
| Daytime sleep episodes, n (%) | 21 (30) | 65 (38) | 88 (34) | 174 (35) | .46 |
| Other, n (%) | 0 (0) | 1 (1) | 4 (2) | 5 (1) | .57[c] |
| No consequences, n (%) | 1 (1) | 1 (1) | 3 (1) | 5 (1) | .84[c] |

[a]ISI: Insomnia Severity Index.

[b]Multiple choice question.

[c]Fisher exact test.



XSL•FO
RenderX

**Table .** Consequences of insomnia symptoms on daily life in EU5[a] (n=730) participants with an ISI[b] score >0**.** Participants were asked "What are/were the consequences of your sleep difficulty symptom(s) on your daily life during your most recent depressive episode?"

| Consequences of insomnia symptoms[c] | EU5 participants with ISI score >0 | | | | |
|---|---|---|---|---|---|
| | Not currently in a depressive episode (n=87) | Unsure but experiencing depression symptoms (n=200) | Currently in a depressive episode (n=443) | Total (N=730) | *P* value |
| Irritability, n (%) | 48 (55) | 134 (67) | 320 (72) | 502 (69) | .006 |
| Reduced productivity, n (%) | 45 (52) | 104 (52) | 226 (51) | 375 (51) | .97 |
| Concentration difficulties, n (%) | 40 (46) | 113 (56) | 272 (61) | 425 (58) | .02 |
| Increased caffeine consumption, n (%) | 36 (41) | 90 (45) | 194 (44) | 320 (44) | .85 |
| Daytime sleep episodes, n (%) | 25 (29) | 49 (24) | 97 (22) | 171 (23 | .35 |
| Other, n (%) | 0 (0) | 3 (1) | 5 (1) | 8 (1) | .77[d] |
| No consequences, n (%) | 5 (6) | 14 (7) | 19 (4) | 38 (5) | .35 |

[a]EU5: France, Germany, Italy, Spain, and the United Kingdom.

[b]ISI: Insomnia Severity Index.

[c]Multiple choice question.

[d]Fisher exact test.

### *Impact of MDDIS Symptoms on Daily Life*

When asked to rank the MDDIS symptoms that had the greatest impact on their lives, most participants in both regions identified depressed mood, lack of motivation, and tiredness as the most disruptive (Figure 5). Among those currently experiencing a depressive episode or symptoms, feelings of worthlessness, loss of energy, and diminished interest were also frequently prioritized.

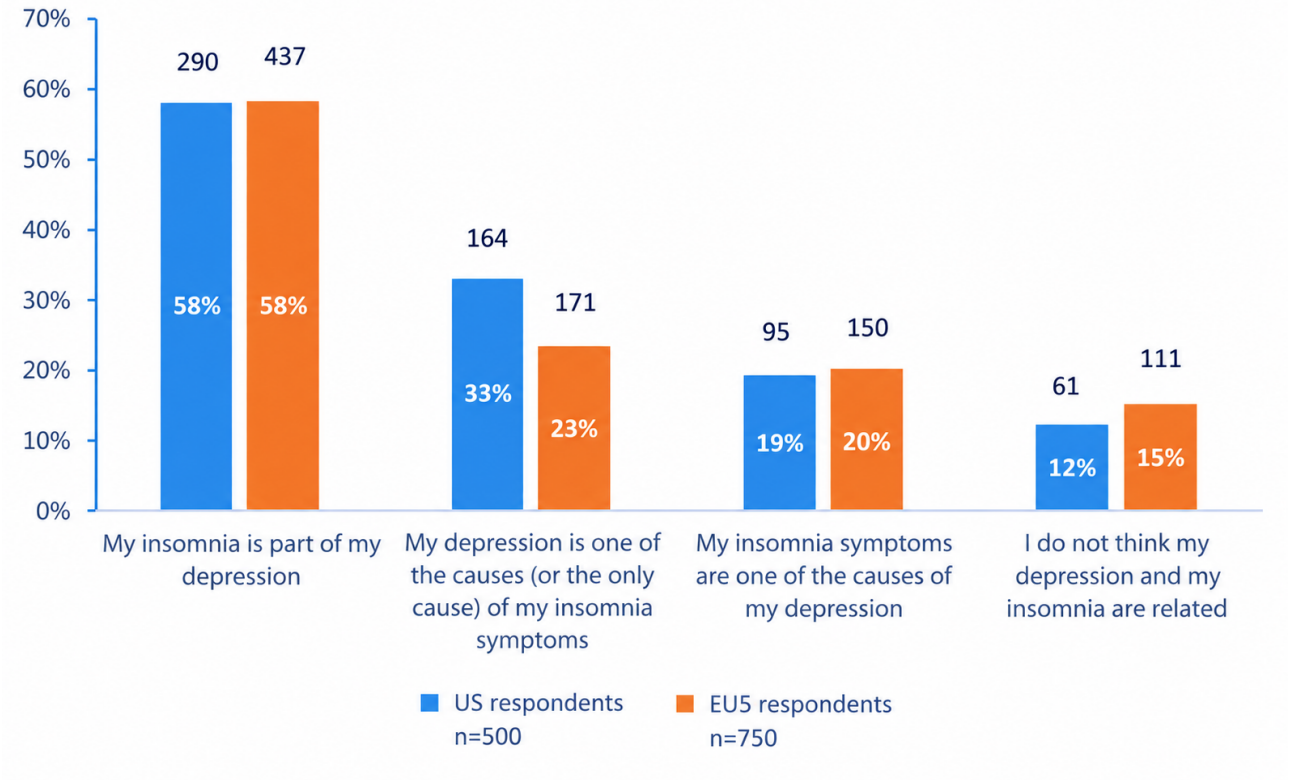
## Perceived Relationship Between Insomnia Symptoms and Depression

When asked how they think insomnia is related to their depression, the majority of participants, 88% (439/500) in the United States and 85% (639/750) in the EU5, believed both conditions are linked. In both regions, 58% (United States: 290/500; EU5: 437/750) considered insomnia symptoms to be part of their depression, while 33% (164/500; United States) and 23% (171/750; EU5) believed their depression was a cause of their insomnia symptoms. Around 20% (United States: 95/500; EU5: 150/750) in each region felt that insomnia symptoms contributed to their depression (Figure 7).



XSL•FO
RenderX

**Figure 7.** Perceived relationship between insomnia symptoms and depression in the US (n=500) and EU5 participants (n=750). EU5: France, Germany, Italy, Spain, and the United Kingdom.




## Satisfaction With Treatment

### *Patient Satisfaction With Antidepressant Treatment for Depressive Symptoms*

Satisfaction with antidepressant effectiveness for depressive symptoms was mixed in both regions. In the United States, 41% (193/472) of participants were dissatisfied, 28% (135/472) were satisfied, and 31% (144/472) reported neutral satisfaction (Table S2 in Multimedia Appendix 2). The distribution of satisfaction ratings differed significantly according to current depressive episode status and depression severity (both $P<.001$) but not according to MDD duration; a greater proportion of patients currently in a depressive episode (121/245, 50%) or experiencing severe depression (93/159, 59%) reported being not satisfied or not satisfied at all. In the EU5 sample, 31% (197/647) of participants reported dissatisfaction, 35% (230/647) reported satisfaction, and 34% (220/647) reported neutral satisfaction. Similarly, satisfaction with antidepressants for depressive symptoms differed significantly according to current depressive episode status and depression severity (both $P<.001$), whereas no significant differences were observed across MDD duration categories. Overall, 40% (157/395) of those currently in a depressive episode and 46% (122/265) of those with severe depression reported being not satisfied or not satisfied at all.

### *Patient Satisfaction With Antidepressant Treatment for Sleep Symptoms*

In the US sample, 57% (269/472) of participants were dissatisfied with the effectiveness of antidepressants for improving sleep, whereas 14% (68/472) reported being satisfied. Satisfaction with antidepressants for sleep symptoms differed significantly according to current depressive episode status ($P=.01$) and depression severity ($P=.006$); 35% (56/159) of participants with severe depression were not satisfied at all (Tables 10 and 11). In the EU5 sample, 41% (264/647) of participants reported being dissatisfied and 26% (168/647) reported being satisfied with the effectiveness of antidepressants for improving sleep. Similarly, satisfaction ratings differed significantly according to current depressive episode status and depression severity (both $P<.001$), with 20% (54/265) of those with severe depression reporting being not satisfied at all. In both regions, no significant differences were observed across MDD duration categories (Tables 10 and 11).



XSL•FO
RenderX

**Table .** Satisfaction with antidepressant effectiveness for sleep problems during most recent depressive episode.

| Satisfaction with antidepressant effectiveness for sleep problems during most recent depressive episode | US population (n=472) | | | | | *P* values |
|---|---|---|---|---|---|---|
| | Not satisfied at all, n (%) | Not satisfied, n (%) | Neutral, n (%) | Satisfied, n (%) | Very satisfied, n (%) | |
| Overall | 121 (26) | 148 (31) | 135 (29) | 58 (12) | 10 (2) | —[a] |
| MDD[b] duration (years) | | | | | | .62 |
| 0 - 5 (n=184) | 44 (24) | 51 (28) | 60 (33) | 24 (13) | 5 (3) | |
| 6 - 20 (n=146) | 35 (24) | 49 (34) | 43 (29) | 17 (12) | 2 (1) | |
| >20 (n=142) | 42 (30) | 48 (34) | 32 (23) | 17 (12) | 3 (2) | |
| Current depression status | | | | | | .01 |
| Not currently in a depressive episode (n=67) | 12 (18) | 13 (19) | 24 (36) | 15 (22) | 3 (4) | |
| Unsure but experiencing depression symptoms (n=160) | 39 (24) | 51 (32) | 45 (28) | 23 (14) | 2 (1) | |
| Currently in a depressive episode (n=245) | 70 (29) | 84 (34) | 66 (27) | 20 (8) | 5 (2) | |
| Depression severity | | | | | | .006[c] |
| None-minimal (n=13) | 2 (15) | 1 (8) | 4 (31) | 5 (38) | 1 (8) | |
| Mild (n=63) | 11 (17) | 18 (29) | 24 (38) | 9 (14) | 1 (2) | |
| Moderate (n=122) | 22 (18) | 38 (31) | 39 (32) | 19 (16) | 4 (3) | |
| Moderately severe (n=115) | 30 (26) | 44 (38) | 30 (26) | 9 (8) | 2 (2) | |
| Severe (n=159) | 56 (35) | 47 (30) | 38 (24) | 16 (10) | 2 (1) | |

[a]Not available.

[b]MDD: major depressive disorder.

[c]Fisher exact test.



XSL•FO
RenderX

**Table .** Satisfaction with antidepressant effectiveness for sleep problems during most recent depressive episode.

| Satisfaction with antidepressant effectiveness for sleep problems | EU5[a] population (n=647) | | | | | *P* values |
|---|---|---|---|---|---|---|
| | Not satisfied at all, n (%) | Not satisfied, n (%) | Neutral, n (%) | Satisfied, n (%) | Very satisfied, n (%) | |
| Overall | 89 (14) | 175 (27) | 215 (33) | 137 (21) | 31 (5) | —[b] |
| MDD[c] duration (years) | | | | | | .19 |
| 0 - 5 (n=278) | 35 (13) | 78 (28) | 92 (33) | 54 (19) | 19 (7) | |
| 6 - 20 (n=213) | 36 (17) | 59 (28) | 71 (33) | 43 (20) | 4 (2) | |
| >20 (n=156) | 18 (12) | 38 (24) | 52 (33) | 40 (26) | 8 (5) | |
| Current depression status | | | | | | <.001 |
| Not currently in a depressive episode (n=84) | 7 (8) | 8 (10) | 27 (32) | 33 (39) | 9 (11) | |
| Unsure but experiencing depression symptoms (n=168) | 21 (13) | 51 (30) | 56 (33) | 33 (20) | 7 (4) | |
| Currently in a depressive episode (n=395) | 61 (15) | 116 (29) | 132 (33) | 71 (18) | 15 (4) | |
| Depression severity | | | | | | <.001[d] |
| None or minimal (n=24) | 1 (4) | 2 (8) | 6 (25) | 5 (21) | 10 (42) | |
| Mild (n=57) | 4 (7) | 6 (11) | 24 (42) | 19 (33) | 4 (7) | |
| Moderate (n=123) | 7 (6) | 29 (24) | 41 (33) | 37 (30) | 9 (7) | |
| Moderately severe (n=178) | 23 (13) | 47 (26) | 69 (39) | 36 (20) | 3 (2) | |
| Severe (n=265) | 54 (20) | 91 (34) | 75 (28) | 40 (15) | 5 (2) | |

[a]EU5: France, Germany, Italy, Spain, and the United Kingdom.

[b]Not available.

[c]MDD: major depressive disorder.

[d]Fisher exact test.

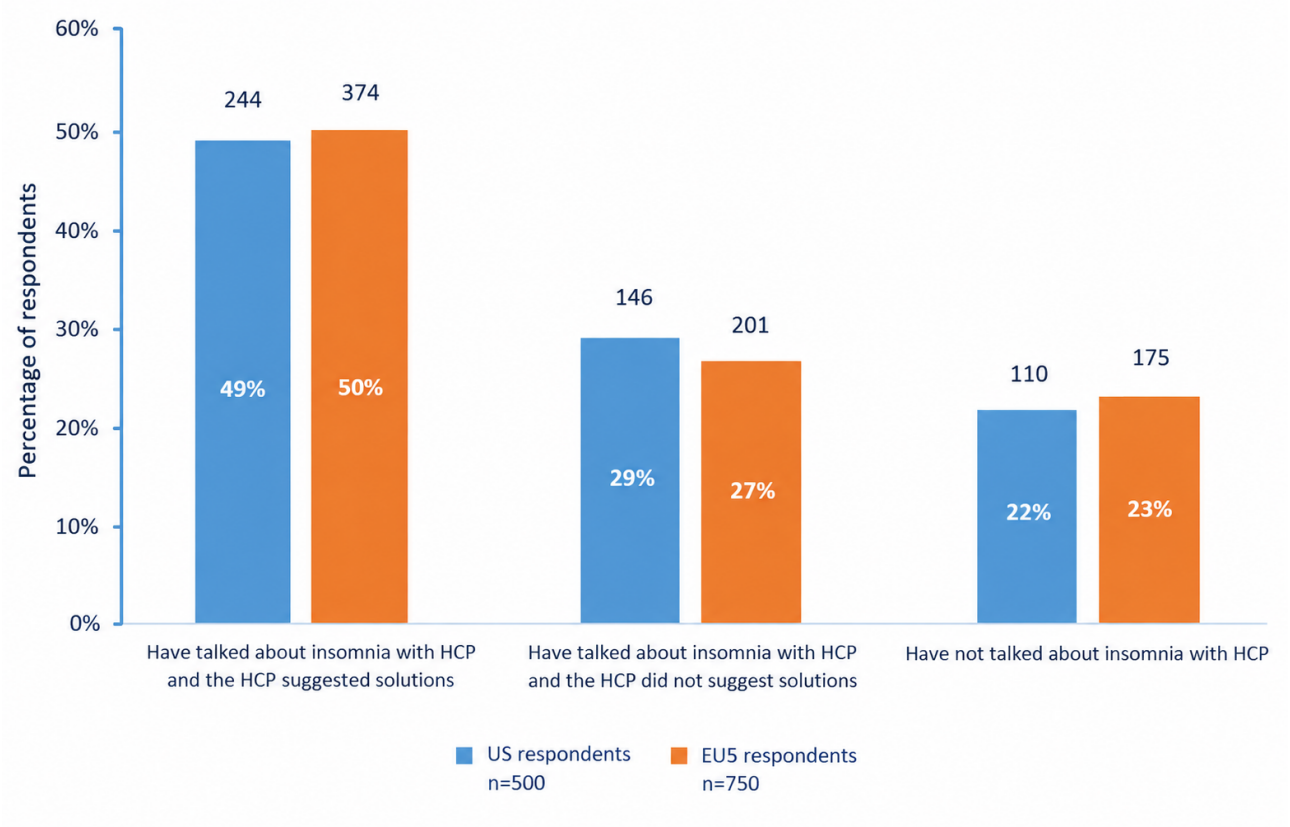
### *Unmet Needs and Desire for Alternative Treatments*

A strong desire for alternative treatments for both depression and sleep difficulties was reported (354/500, 71% in the United States and 488/750, 65% in the EU5), particularly among participants experiencing a current depressive episode or with more severe depression and insomnia symptoms. While most participants had discussed their insomnia symptoms with a clinician or another member of their care team (United States: 390/500 and 78%, EU5: 575/750, 77%), over one-third (United States: 146/390, 37% and EU5: 201/575, 35%) reported receiving no recommendations, especially those with more severe insomnia symptoms (Figure 8).



XSL•FO
RenderX

**Figure 8.** Discussion related to insomnia with health care professionals in the United States (n=500) and EU5 participants (n=750). EU5: France, Germany, Italy, Spain, and the United Kingdom; HCP: health care professional.




## Discussion

### Main Findings

This cross-sectional study provides patient-reported evidence for the burden of insomnia symptoms among individuals with MDD. Overall, insomnia symptoms were commonly reported despite treatment use, with moderate to severe symptoms more commonly reported among participants with more severe depression and those currently experiencing a depressive episode. Participants also described a substantial impact of insomnia symptoms on daily functioning, emotional well-being, cognitive functioning, and productivity, expressing low satisfaction with the perceived effectiveness of current treatments for sleep symptoms and a desire for additional therapeutic options. Collectively, these findings highlight the considerable patient-reported burden associated with insomnia symptoms and suggest that persistent sleep difficulties remain an important aspect of the patient experience.

This study shows that insomnia symptoms remain highly prevalent among patients with MDD, even among individuals receiving treatment. These findings are consistent with previous epidemiological studies showing that sleep disturbance is a common feature of depression and that insomnia symptoms may persist despite improvement in depressive symptoms, increasing the risk of relapse and recurrence [14,15,30,31]. Furthermore, the results suggest a link between insomnia severity and depression severity; participants with more severe depressive symptoms were more likely to report greater insomnia severity, in line with previous literature showing that short sleep and insomnia are independently associated with increased depression severity and symptom frequency [39,40]. These findings support the importance of considering sleep disturbance as a clinically relevant component of the overall burden of MDD, in particular among those with more severe depression.

Beyond their high prevalence, insomnia symptoms were perceived as having a substantial impact on multiple aspects of daily functioning. Participants reported impairments affecting emotional well-being, cognitive functioning, or work productivity, consistent with previous studies demonstrating that people with both depression and sleep problems experience daytime symptoms such as poor concentration and fatigue, and that insomnia symptoms have a strong negative impact on QoL and psychological distress [15,17].

Participants' perceptions further illustrate the complexity of the relationship between insomnia symptoms and depression. Although insomnia is recognized as a core symptom of MDD [14], some participants still do not perceive their sleep difficulties as being part of their depression. This finding may reflect limited awareness of the relationship between these conditions and supports the need to provide patients with information regarding the coexistence of mood and sleep symptoms [14,16,30].

Treatment use was high, despite the prevalence of insomnia symptoms. The majority of participants reported dissatisfaction with the perceived effectiveness of antidepressants for improving sleep symptoms and expressed interest in alternative therapeutic options addressing both depression and sleep. Previous literature



XSL•FO
RenderX

has observed that more than half of patients perceived antidepressants as insufficient for improving sleep and seek additional support [15]. However, this study did not assess treatment adherence, treatment duration, clinical appropriateness, or treatment resistance. Therefore, these findings should be interpreted as reflecting patients' perceptions of persistent sleep difficulties despite treatment rather than evidence of inadequate clinical management.

Despite the well-established literature demonstrating an association between depression and insomnia symptoms [14,16,30-32,41], this association is not adequately reflected in current treatment practices. Although most participants reported discussing their sleep difficulties with a clinician or another member of their care team, over one-third indicated that no treatment suggestions were provided. As the study did not capture the reasons underlying these interactions, alternative explanations, including patient preferences, clinical judgment, or treatment prioritization, cannot be excluded. Nevertheless, these findings suggest a potential communication or care gap that warrants further investigation.

### Strengths and Limitations

A major strength of this study is its large multinational sample recruited through an online patient community, allowing the collection of patient-reported experiences from individuals with MDD across 6 countries [42]. The use of validated instruments (PHQ-9 and ISI) ensured standardized assessment of depression and insomnia severity, while the study-specific questionnaire provided valuable insights into patients' treatment experiences, perceptions, and unmet needs. Compared with previous studies that primarily focused on the economic burden and health care resource use associated with insomnia symptoms in MDD [31,32], this study offers a broader understanding of the patient experience and highlights aspects of disease burden that are not routinely captured in clinical practice.

Several limitations should nevertheless be acknowledged. Recruitment through an online platform may have introduced selection bias by excluding individuals without internet access or those less familiar with digital tools, and by preferentially including patients who are more engaged in managing their condition. In addition, the sample was predominantly composed of cisgender women, which may limit the generalizability of the findings to the broader MDD population [7]. The observed differences in the proportion of participants identifying as transgender or nonbinary between the United States and the EU5 may also reflect cultural and societal differences in the willingness to disclose sex identity.

All data were self-reported, introducing the possibility of recall and reporting bias. In particular, the diagnosis of MDD was based on participant self-report and was not independently confirmed through medical records or health care professionals, which may have resulted in diagnostic misclassification. Furthermore, information regarding treatment adherence, treatment duration, treatment optimization, clinician rationale for treatment selection, and treatment resistance was not collected, limiting the interpretation of participants' reported treatment experiences. Sleep duration, a factor known to influence the relationship between insomnia and depression, was also not assessed and may have acted as a confounder. Finally, the cross-sectional design precludes any conclusions regarding temporal or causal relationships between insomnia symptoms and depression severity.

Future longitudinal studies integrating both patient- and clinician-reported outcomes are warranted to better characterize the relationship between insomnia symptoms and the onset, course, and treatment response of MDD. Such studies should also investigate behavioral and educational aspects of insomnia management, including patients' understanding of the relationship between sleep and depression, as well as strategies that promote adaptive coping and improve long-term outcomes.

### Implications for Clinical Practice

Given the well-established association between depression and insomnia symptoms reported in the literature [38], our findings further emphasize the importance of considering sleep symptoms during the clinical assessment of patients with MDD. Although this study cannot determine the appropriateness of current treatment strategies, the persistence of patient-reported sleep difficulties and the expressed need for additional therapeutic options suggest that insomnia symptoms remain an important component of patient care deserving continued clinical attention.

## Conclusion

This multinational patient study highlights the substantial burden associated with insomnia symptoms among individuals with MDD. Participants reported an impact of insomnia symptoms on daily functioning, emotional well-being, and productivity. More severe depression was accompanied by a greater proportion of participants with moderate or severe insomnia symptoms, and dissatisfaction with current treatment options was widespread. Around two-thirds of participants also expressed a desire for additional treatment options. Although the cross-sectional design and patient-reported nature of the data preclude conclusions regarding causality or the appropriateness of current clinical management, these findings suggest that insomnia symptoms remain an important component of the patient experience in MDD and identify a potential care gap that warrants further investigation.

## Funding

This work was funded by Johnson & Johnson Innovative Medicine.

## Authors' Contributions

Conceptualization: ZZ, EP, JR, JH



XSL•FO
RenderX

Data curation: EP
Formal analysis: EP, ZZ, JR, JH
Funding acquisition: ZZ
Investigation: EP, ZZ, JR, JH
Methodology: ZZ, EP, JR, JH
Project administration: ZZ, EP
Resources: EP, ZZ
Software: EP
Supervision: ZZ (lead), EP (equal)
Validation: EP (lead), ZZ (equal), JR (supporting), JH (supporting)
Visualization: EP (lead), ZZ (equal), JR (supporting), JH (supporting)
Writing – original draft: EP (lead), ZZ (equal), JR (supporting), JH (supporting)
Writing – review & editing: EP (lead), ZZ (equal), JR (supporting), JH (supporting)

## Conflicts of Interest

EP is an employee of Carenity. ZZ and JR are employees of Johnson and Johnson Innovative Medicine. JH is a psychiatric nurse practitioner and is a consulting speaker for Tempus AI, Inc., a consultant for the Point of Care Network, and a consultant for Johnson & Johnson Innovative Medicine. No other conflicts of interest were declared.

Multimedia Appendix 1
Characteristics of EU5 study population per country.
[PDF File, 168 KB - i-jmr_v15i1e93231_app1.pdf ]

Multimedia Appendix 2
Satisfaction with antidepressant effectiveness in treating depression during most recent depressive episode (US: n=472, EU5: n=647).
[PDF File, 132 KB - i-jmr_v15i1e93231_app2.pdf ]

## References

1. What is depression? Anxiety and Depression Association of America. URL: https://adaa.org/understanding-anxiety/depression#Types%20of%20Depression [accessed 2026-08-12]
2. Malhi GS, Mann JJ. Depression. Lancet 2018 Nov 24;392(10161):2299-2312. [doi: 10.1016/S0140-6736(18)31948-2] [Medline: 30396512]
3. Arnaud AM, Brister TS, Duckworth K, et al. Impact of major depressive disorder on comorbidities: a systematic literature review. J Clin Psychiatry 2022 Oct 19;83(6):21r14328. [doi: 10.4088/JCP.21r14328] [Medline: 36264099]
4. Otte C, Gold SM, Penninx BW, et al. Major depressive disorder. Nat Rev Dis Primers 2016 Sep 15;2:16065. [doi: 10.1038/nrdp.2016.65] [Medline: 27629598]
5. APA dictionary of psychology. American Psychological Association. URL: https://dictionary.apa.org/major-depressive-episode [accessed 2026-08-12]
6. Hasin DS, Sarvet AL, Meyers JL, et al. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. JAMA Psychiatry 2018 Apr 1;75(4):336-346. [doi: 10.1001/jamapsychiatry.2017.4602] [Medline: 29450462]
7. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2021 (GBD 2021): Institute for Health Metrics and Evaluation (IHME); 2024. URL: https://vizhub.healthdata.org/gbd-results/ [accessed 2025-04-01]
8. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov 6;398(10312):1700-1712. [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
9. GBD 2021 Diseases and Injuries Collaborators. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2024 May 18;403(10440):2133-2161. [doi: 10.1016/S0140-6736(24)00757-8] [Medline: 38642570]
10. Simon J, Wienand D, Park AL, et al. Excess resource use and costs of physical comorbidities in individuals with mental health disorders: a systematic literature review and meta-analysis. Eur Neuropsychopharmacol 2023 Jan;66:14-27. [doi: 10.1016/j.euroneuro.2022.10.001] [Medline: 36345094]
11. Ishak WW, Balayan K, Bresee C, et al. A descriptive analysis of quality of life using patient-reported measures in major depressive disorder in a naturalistic outpatient setting. Qual Life Res 2013 Apr;22(3):585-596. [doi: 10.1007/s11136-012-0187-6] [Medline: 22544416]
12. Daly EJ, Trivedi MH, Wisniewski SR, et al. Health-related quality of life in depression: a STAR*D report. Ann Clin Psychiatry 2010 Feb;22(1):43-55. [Medline: 20196982]



XSL•FO
RenderX

13. Berlim MT, Fleck MP. Quality of life and major depression. In: Ritsner MS, Awad AG, editors. Quality of Life Impairment in Schizophrenia, Mood and Anxiety Disorders: Springer; 2007. [doi: 10.1007/978-1-4020-5779-3_12]
14. Sunderajan P, Gaynes BN, Wisniewski SR, et al. Insomnia in patients with depression: a STAR*D report. CNS Spectr 2010 Jun;15(6):394-404. [doi: 10.1017/s1092852900029266] [Medline: 20625372]
15. Paterson LM, Nutt DJ, Wilson SJ. NAPSAQ-1: National Patient Sleep Assessment Questionnaire in depression. Int J Psychiatry Clin Pract 2009;13(1):48-58. [doi: 10.1080/13651500802450498] [Medline: 24946122]
16. Fang H, Tu S, Sheng J, Shao A. Depression in sleep disturbance: a review on a bidirectional relationship, mechanisms and treatment. J Cell Mol Med 2019 Apr;23(4):2324-2332. [doi: 10.1111/jcmm.14170] [Medline: 30734486]
17. Katz DA, McHorney CA. The relationship between insomnia and health-related quality of life in patients with chronic illness. J Fam Pract 2002 Mar;51(3):229-235. [Medline: 11978233]
18. Murphy MJ, Peterson MJ. Sleep Disturbances in Depression. Sleep Med Clin 2015 Mar;10(1):17-23. [doi: 10.1016/j.jsmc.2014.11.009] [Medline: 26055669]
19. Thase ME. Depression and sleep: pathophysiology and treatment. Dialogues Clin Neurosci 2006;8(2):217-226. [doi: 10.31887/DCNS.2006.8.2/mthase] [Medline: 16889107]
20. Riemann D, Voderholzer U. Primary insomnia: a risk factor to develop depression? J Affect Disord 2003 Sep;76(1-3):255-259. [doi: 10.1016/s0165-0327(02)00072-1] [Medline: 12943956]
21. Nierenberg AA, Husain MM, Trivedi MH, et al. Residual symptoms after remission of major depressive disorder with citalopram and risk of relapse: a STAR*D report. Psychol Med 2010 Jan;40(1):41-50. [doi: 10.1017/S0033291709006011] [Medline: 19460188]
22. Buysse DJ, Kupfer DJ, Cherry C, Stapf D, Frank E. Effects of prior fluoxetine treatment on EEG sleep in women with recurrent depression. Neuropsychopharmacology 1999 Aug;21(2):258-267. [doi: 10.1016/S0893-133X(99)00025-1] [Medline: 10432474]
23. Thase ME, Simons AD, Reynolds CF. Abnormal electroencephalographic sleep profiles in major depression: association with response to cognitive behavior therapy. Arch Gen Psychiatry 1996 Feb;53(2):99-108. [doi: 10.1001/archpsyc.1996.01830020013003] [Medline: 8629894]
24. McCall WV, Black CG. The link between suicide and insomnia: theoretical mechanisms. Curr Psychiatry Rep 2013 Sep;15(9):389. [doi: 10.1007/s11920-013-0389-9] [Medline: 23949486]
25. Woznica AA, Carney CE, Kuo JR, Moss TG. The insomnia and suicide link: toward an enhanced understanding of this relationship. Sleep Med Rev 2015 Aug;22:37-46. [doi: 10.1016/j.smrv.2014.10.004] [Medline: 25454672]
26. O'Connor SJ, Hewitt N, Kuc J, Orsini LS. Predictors and risk factors of treatment-resistant depression. J Clin Psychiatry 2024;85(1):23r14885. [doi: 10.4088/JCP.23r14885]
27. Brietzke E, Vazquez GH, Kang MJY, Soares CN. Pharmacological treatment for insomnia in patients with major depressive disorder. Expert Opin Pharmacother 2019 Aug;20(11):1341-1349. [doi: 10.1080/14656566.2019.1614562] [Medline: 31046480]
28. Carney CE, Segal ZV, Edinger JD, Krystal AD. A comparison of rates of residual insomnia symptoms following pharmacotherapy or cognitive-behavioral therapy for major depressive disorder. J Clin Psychiatry 2007 Feb;68(2):254-260. [doi: 10.4088/jcp.v68n0211] [Medline: 17335324]
29. Carney CE, Harris AL, Friedman J, Segal ZV. Residual sleep beliefs and sleep disturbance following cognitive behavioral therapy for major depression. Depress Anxiety 2011 Jun;28(6):464-470. [doi: 10.1002/da.20811] [Medline: 21400641]
30. Franzen PL, Buysse DJ. Sleep disturbances and depression: risk relationships for subsequent depression and therapeutic implications. Dialogues Clin Neurosci 2008;10(4):473-481. [doi: 10.31887/DCNS.2008.10.4/plfranzen] [Medline: 19170404]
31. Joshi K, Cambron-Mellott MJ, Costantino H, Pfau A, Jha MK. The real-world burden of adults with major depressive disorder with moderate or severe insomnia symptoms in the United States. J Affect Disord 2023 Feb 15;323:698-706. [doi: 10.1016/j.jad.2022.12.005] [Medline: 36481229]
32. Riise JG, Cambron-Mellott MJ, Zhang Z, et al. Examining the burden of major depressive disorder with moderate-to-severe insomnia symptoms in five European countries. J Affect Disord 2025 Oct 1;386:119403. [doi: 10.1016/j.jad.2025.119403] [Medline: 40378968]
33. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
34. Gómez-Gómez I, Benítez I, Bellón J, et al. Utility of PHQ-2, PHQ-8 and PHQ-9 for detecting major depression in primary health care: a validation study in Spain. Psychol Med 2023 Sep;53(12):5625-5635. [doi: 10.1017/S0033291722002835] [Medline: 36258639]
35. Arthurs E, Steele RJ, Hudson M, Baron M, Thombs BD, (CSRG) Canadian Scleroderma Research Group. Are scores on English and French versions of the PHQ-9 comparable? an assessment of differential item functioning. PLoS One 2012;7(12):e52028. [doi: 10.1371/journal.pone.0052028] [Medline: 23251676]
36. Kliem S, Sachser C, Lohmann A, et al. Psychometric evaluation and community norms of the PHQ-9, based on a representative German sample. Front Psychiatry 2024;15:1483782. [doi: 10.3389/fpsyt.2024.1483782] [Medline: 39726913]




XSL•FO
RenderX

37. Di Matteo R, Bolgeo T, Simonelli N, et al. Psychometric properties and measurement invariance of the Patient Health Questionnaire 9 in an Italian coronary heart disease population. J Cardiovasc Nurs 2025 Feb 10. [doi: 10.1097/JCN.0000000000001178] [Medline: 39932691]
38. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]
39. Tubbs AS, Gallagher R, Perlis ML, et al. Relationship between insomnia and depression in a community sample depends on habitual sleep duration. Sleep Biol Rhythms 2020 Apr;18(2):143-153. [doi: 10.1007/s41105-020-00255-z] [Medline: 34305449]
40. Tu Y, Qing G, Chen M, Chen H. Association between depression severity and trouble sleeping: a population-based study. Medicine (Baltimore) 2024 Sep 6;103(36):e39611. [doi: 10.1097/MD.0000000000039611] [Medline: 39252292]
41. Sivertsen B, Salo P, Mykletun A, et al. The bidirectional association between depression and insomnia: the HUNT study. Psychosom Med 2012 Sep;74(7):758-765. [doi: 10.1097/PSY.0b013e3182648619] [Medline: 22879427]
42. Wicks P, Stamford J, Grootenhuis MA, Haverman L, Ahmed S. Innovations in e-health. Qual Life Res 2014 Feb;23(1):195-203. [doi: 10.1007/s11136-013-0458-x] [Medline: 23852096]

## Abbreviations

***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition)
**EU5:** France, Germany, Italy, Spain, and the United Kingdom
**GBD:** Global Burden of Disease
**ISI:** Insomnia Severity Index
**MDD:** major depressive disorder
**MDDIS:** MDD experiencing insomnia symptoms
**OTC:** over-the-counter
**PHQ-9:** Patient Health Questionnaire-9
**QoL:** quality of life

*Edited by M Balcarras; submitted 10.Feb.2026; peer-reviewed by D Paudel, L Tamam; revised version received 24.Jul.2026; accepted 24.Jul.2026; published 21.Sep.2026.*

*Please cite as:*
*Hamilton J, Pain E, Riise J, Zhang Z*
*Burden of Insomnia Among Individuals With Major Depressive Disorder: Cross-Sectional Observational Study*
*Interact J Med Res 2026;15:e93231*
*URL: https://www.i-jmr.org/2026/1/e93231*
*doi:10.2196/93231*

© Josh Hamilton, Emilie Pain, Jesper Riise, Zhiheng Zhang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 21.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Patients With Acute Stroke Symptoms Consulting Large Language Model Chatbots Before Presenting to the Emergency Department: Case Series

Sara Dierickx[1,2*], MD; Paolo Speziali[3*], MSc; Helena Gossye[1,2*], MD, PhD; Sofia Maldonado Slootjes[1,2*], MD; Sylvie De Raedt[1,2*], MD, PHD; Anaïs Missiaen[1,2*], MD; Michel Van Lint[4*], MD; Edwin Van Dellen[5*], MD, PHD; Fenne Vandervorst[1,2*], MD

[1]NEUR Research Group, Center for Neurosciences (C4N), Vrije Universiteit Brussel, Brussels, Belgium

[2]Department of Neurology, Universitair Ziekenhuis Brussel, Laarbeeklaan 101, Brussels, Belgium

[3]AI Research Group, Vrije Universiteit Brussel, Brussels, Belgium

[4]Department of Ophthalmology, Universitair Ziekenhuis Brussel, Brussels, Belgium

[5]Department of Psychiatry, Brain Centre, University Medical Center Utrecht, Utrecht, The Netherlands

[*]all authors contributed equally

**Corresponding Author:**
Fenne Vandervorst, MD
NEUR Research Group, Center for Neurosciences (C4N), Vrije Universiteit Brussel, Brussels, Belgium

## Abstract

**Background:** Chatbots are being increasingly consulted for health-related questions. Understanding their potential to recognize serious conditions such as stroke is crucial for assessing their role in clinical decision-making and public health.

**Objective:** This case series describes 3 patients who consulted large language model (LLM)–powered chatbots before seeking care at the emergency department and explores the potential benefits and limitations of such chatbot use in the context of acute stroke symptoms.

**Methods:** Cases were identified when patients or their family members spontaneously reported during history-taking that they had consulted a chatbot at symptom onset and the corresponding chat history was available for review. We describe 3 patients who consulted LLM-powered chatbots before presenting to the emergency department with symptoms subsequently diagnosed as stroke. The chatbot interactions were reviewed for recognition of stroke as a possible diagnosis and for recommendations to seek medical care.

**Results:** In 2 cases, the LLM-powered chatbots correctly recognized the likelihood of stroke and the need for urgent medical assessment. In the third case, the symptoms were not initially recognized as signs of stroke, and medical help was advised only once they persisted.

**Conclusions:** LLM-powered chatbots have the potential to recognize stroke symptoms and promote timely access to medical care. However, their responses may not reliably identify stroke symptoms or provide sufficiently urgent advice. Deployment of such chatbots in the context of acute neurological symptoms requires clinical oversight and patient education regarding their limitations.

***(Interact J Med Res 2026;15:e107318)*** doi:10.2196/107318

**KEYWORDS**

artificial intelligence; AI; ischemic stroke; transient; large language models; public health; stroke

## Background

Stroke symptoms are highly variable and depend largely on the location of the vascular lesion within the central nervous system. A defining feature of symptoms caused by cerebrovascular events is their sudden onset, which helps distinguish them from other central nervous system disorders, such as demyelinating diseases. The FAST (face drooping, arm weakness, speech difficulties, time to call emergency services) mnemonic summarizes key stroke symptoms, with the “time” component emphasizing the urgency of immediate medical attention [1]. Timely treatment—including thrombolysis and thrombectomy—remains the most critical determinant of outcome, with every minute being decisive. Mass media campaigns focusing on individual components of the FAST acronym have been shown to improve public awareness of stroke symptoms and increase thrombolysis rates [2]. However, their effects tend to diminish after approximately 6 months,



XSL•FO
RenderX

highlighting the need for repeated initiatives. Additionally, the FAST acronym is less sensitive in detecting posterior circulation stroke symptoms, often presenting with ataxia and visual changes. More extensive acronyms have been developed and tested; however, they were associated with lower retention rates [3].

AI systems such as ChatGPT 4.1 (OpenAI) [4] and Grok 3 (SpaceXSI) [5] are becoming increasingly accessible worldwide. Large language model (LLM)–based systems can extract information from conversations, integrate relevant context, and apply advanced communication skills. In the setting of acute stroke, history-taking can be exceptionally challenging. Patients may experience anosognosia, visual or language disturbances such as hemianopsia and aphasia, or impaired use of an upper limb due to paresis, and/or neglect, all of which can limit their ability to provide accurate information or interact effectively.

At our hospital, 3 patients experiencing acute ischemic stroke presented at the emergency department after consulting an LLM-powered AI chatbot. Atypical stroke symptoms were correctly identified in one conversation, despite—or possibly due to—the presence of spelling errors. The original and translated conversations are provided in Multimedia Appendices 1-3.

Of 3 patients, 2 were appropriately prompted to seek emergency medical care. The third patient was initially advised to rest; however, after 13 hours of persistent symptoms, he was referred to seek medical attention. Each case is described separately, followed by a brief interpretation and commentary by medical specialists.

These cases exemplify an emerging phenomenon increasingly encountered in clinical practice: patients using AI chatbots to seek medical advice regarding symptoms they experience. Given the growing accessibility and use of these technologies, it is important to understand their potential role in the diagnostic pathway. The aim of this article is to evaluate the benefits and risks associated with chatbot-based consultations by linking chatbot interactions with clinical and radiological findings. In the Discussion section, we further explore the practical implications of these findings and consider future perspectives on the integration of chatbots into health care systems.

## Case Presentation

### Ethical Considerations

Ethics approval was not required for this case report. Written informed consent was obtained from all patients included in this case report.

### Case 1

#### *Description*

A 79-year-old male patient with a medical history of diabetes and hypertension experienced sudden blurred vision lasting approximately 4 to 5 minutes. One hour later, he experienced transient word-finding difficulties for 3 to 4 minutes. He subsequently drove a 54-year-old female friend to a dental appointment and mentioned these episodes to her. Curious about the potential causes of such speech difficulties, she consulted ChatGPT 4.1 for insights.

ChatGPT 4.1 generated a list of differential diagnoses, including transient ischemic attack (TIA), migraine with aura, focal seizures, multiple sclerosis, stress or fatigue, hypoglycemia, medication side effects, alcohol or drug use, and dry mouth and jaw or dental issues. It advised that if the symptoms were isolated and the individual felt well afterward, the episode could be considered benign. However, if the symptoms recurred and were accompanied by weakness, vision problems, confusion, or numbness, it recommended seeking medical attention. The chatbot then asked whether she wanted to continue the discussion and offered to ask additional questions, to which she agreed. ChatGPT 4.1 then asked about the duration of the symptoms and any contextual factors (eg, stress, fatigue, alcohol consumption, or the possibility of hypoglycemia). It also inquired about associated symptoms, previous episodes, and the patient's medical history. The patient's friend responded that the symptoms lasted a few minutes and occurred without any specific triggers. She mentioned that the patient had experienced blurred vision 1 hour prior, indicating it was the first such episode and that he was being treated for diabetes with metformin and dulaglutide.

On the basis of this exchange, ChatGPT 4.1 suggested that a TIA was the most likely diagnosis and warned her about its association with an elevated stroke risk. It recommended seeking urgent medical attention by going to the emergency department. Furthermore, it provided guidance on how to clearly summarize the events for the physician and explained what to expect during further medical examinations (Multimedia Appendix 1).

The patient subsequently presented to our emergency department and was admitted for a clinical suspicion of recurrent TIA. Computed tomography (CT) scan demonstrated luminal narrowing of the left internal carotid artery. Magnetic resonance imaging (MRI) revealed no evidence of recent cerebral ischemia but confirmed the presence of left internal carotid artery narrowing attributable to an underlying dissection. A diagnosis of retinal and MRI-negative cerebral TIA was considered.

#### *Interpretation*

On the basis of a brief but accurate description of symptoms, the chatbot concluded transient word-finding difficulties. A broad but largely accurate differential diagnosis was given (only focal seizures seem less probable based on the description of dysarthria instead of aphasia). Initially, a more benign phenomenon was suggested; however, no further explanation was provided. The question regarding other neurological deficits in this case appears to have prompted the recommendation to seek urgent medical attention, even though isolated, acute, and transient neurological deficits alone should always warrant medical evaluation.

### Case 2

#### *Description*

A 59-year-old patient with a history of obstructive sleep apnea experienced sudden onset of visual disturbances affecting the



XSL•FO
RenderX

right visual field, along with left-sided neck pain; lightheadedness; and paresthesia in the right arm, leg, and face.

Before seeking medical attention, he consulted the AI chatbot Grok 3, which suggested both urgent and nonurgent explanations, including migraine with aura, TIA, stroke, peripheral neuropathy, nerve compression, hypoglycemia, hyperventilation, and ocular disorders. It advised calling the emergency services (112) if the symptoms were sudden in onset; one-sided; or associated with confusion, weakness, or speech difficulties (Multimedia Appendix 2). As the patient's symptoms were both acute and lateralized, he promptly presented to the emergency department.

A perfusion brain CT scan showed hypoperfusion in the territory of the left posterior cerebral artery, while the noncontrast CT showed no abnormalities, and an occlusion of the P1 segment was found on CT angiography. We administered alteplase and conducted intra-arterial digital subtraction angiography, which revealed that the clot had already dissolved. The patient's symptoms improved within minutes to hours, except for a persistent right-sided inferior quadrantanopia. He was admitted to the neurology ward for further stroke workup. Brain MRI confirmed an acute infarction in the left occipital lobe; the image can be found in Figure S1 in Multimedia Appendix 4.

#### *Interpretation*

In the original description of symptoms, some spelling errors were made, possibly due to the stroke itself causing sensory disturbances and hemianopsia. Nevertheless, symptoms were recognized correctly. Once again, a broad differential diagnosis was listed, with, at the end, instructions about the degree of urgency. Symptoms with an acute onset; unilateral distribution; and the co-occurrence of speech difficulties, confusion, or weakness were correctly identified as requiring urgent medical advice. However, a milder and more fluctuating or recurrent course can also occur in the context of stroke and should not always be considered as nonurgent.

### Case 3

#### *Description*

A 25-year-old male patient, without relevant medical history, experienced weakness in the right arm and right leg, accompanied by speech difficulties and peripheral visual loss. The symptoms started while he was washing his hands after using the toilet. He consulted ChatGPT 4.1 for advice and explained that his head felt "weird" after going to the toilet, while his eyes appeared "crooked." ChatGPT suggested several possible explanations, including a drop in blood pressure, dehydration, hypoglycemia, and excessive straining. It advised seeking medical attention if symptoms such as visual disturbances, weakness, confusion, nausea, vomiting, or loss of balance occurred. The chatbot subsequently asked several follow-up questions. The patient reported he remained dizzy and that his visual disturbances were unchanged; no other symptoms were present. On the basis of this response, ChatGPT concluded that a temporary drop in blood pressure most likely caused the symptoms and advised lying down, drinking water, and breathing slowly. It stated that the symptoms should improve within 5 to 10 minutes and recommended seeking urgent medical attention if they persisted. When subsequently asked whether the symptoms had improved, the patient replied that they had not. ChatGPT then asked whether he experienced tingling or weakness in an arm or leg, which the patient denied. After several additional questions, the patient suggested going to sleep for a while. ChatGPT indicated that this was reasonable but advised informing someone beforehand. The patient also mentioned that he normally had mild strabismus but that it had worsened during this episode. ChatGPT responded that this could be associated with a drop in blood pressure. Later, the patient reported feeling better when lying down, which ChatGPT interpreted as further supporting this explanation. When the patient asked whether the symptoms could be related to a very spicy sauce he had eaten for the first time, ChatGPT indicated that this could potentially contribute to a drop in blood pressure. Finally, when the patient reported that the symptoms persisted after 13 hours, ChatGPT advised him to seek medical attention at the emergency department (Multimedia Appendix 3). At the emergency department, neurological examination revealed dysarthria and pronation of the right arm without drift. No objective motor deficit of the leg could be demonstrated, although the patient reported a subjective difference while walking. A contrast-enhanced CT scan of the head showed no abnormalities. The patient was hospitalized at the Stroke Unit. Subsequently, the brain MRI confirmed an acute ischemic stroke located in the left thalamus. An etiological workup did not reveal an underlying cause.

#### *Interpretation*

In this case, symptoms were described concisely with attention to the circumstances in which they occurred. Stroke was unfortunately not mentioned in the list of possible causes of dizziness and altered vision. The following questions focused on persistence of symptoms as well as associated symptoms. Subsequently, recommendations regarding time frames were not applied consistently by the chatbot, resulting in a delayed presentation to the emergency department. Alternative explanations proposed by the patient, including strabismus and the intake of spicy harissa sauce, were endorsed by the chatbot, which may reflect a tendency of LLMs to reinforce user satisfaction.

## Discussion

### Principal Findings

On the basis of these 3 examples of patients experiencing acute ischemic stroke consulting a LLM-based AI chatbot for medical advice, the following observations were made: first, atypical stroke symptoms can be detected by AI chatbots with little interference from writing or language errors, often occurring in the setting of stroke, due to neurological deficits such as aphasia, hemianopia, and motor impairment; second, all patients described followed the advice given by the chatbot; and third, the way in which certain symptoms are described, including contextual detail and more elaborate phrasing, inherently shaped by the patient's own interpretation of their symptoms, seems to influence how the symptoms are interpreted by the chatbot.



XSL•FO
RenderX

## Opportunities of AI Chatbots

Stroke symptoms remain poorly recognized in the general population, and LLM-powered AI chatbots offer the advantage of being available at all times. In 2 of 3 described cases, they were able to recognize stroke symptoms even when described in lay terms with spelling errors in the second case. Language disorders are a characteristic feature of numerous neuropsychiatric conditions, and LLMs have the capability to detect these disturbances, thereby not only preserving comprehension of the communicated message but also possibly contributing to both diagnosis and prognosis [6,7]. In a more general medical context, the potential of chatbots for safe and effective use has been widely reported [8]. For instance, ChatGPT 4.1 achieved passing grades on the United States Medical Licensing Examination [9]. Moreover, in a cross-sectional study of patient questions in an online forum, the quality and empathy of ChatGPT 4.1 responses were rated significantly higher for both quality and empathy, even when compared with the longest physician-authored responses [10]. More recently, specialized LLM-based AI bots have been created with improved capabilities for clinical reasoning, history-taking, and high-quality conversation [11]. Although the versatility of widely used AI LLMs may yield impressive results, challenges arise in highly specialized domains, such as neurology.

## Risks of AI Chatbots

Concerns have been raised regarding the limitations of medical advice provided by AI chatbots. For example, these systems can make errors due to discrimination or stereotyping [12]. Because the LLMs are created based on unchecked, potentially biased data, these biases may be adopted in their output [13]. Aside from general reliability issues, chatbots that are optimized directly for user satisfaction can develop manipulative strategies aimed at securing positive feedback (also called "sycophancy") [14] and user retention. As a result, they may be more prone to reinforcing user-suggested explanations and prolonging patient interactions rather than directing users toward appropriate clinical evaluation, as seen in case 3. However, the phenomenon of sycophancy also implies a sensitivity of the chatbot to factors such as the patient's experience of certain symptoms, concerns about illness or symptoms, and general well-being—elements that are important in human patient interaction. Nonetheless, this attentiveness becomes problematic when it causes the chatbot to prioritize the patient's own framing over clinically relevant warning symptoms, as appeared to occur in the third case. These findings reinforce the need for human oversight and robust evaluation prior to clinical deployment.

In addition, evaluating a chatbot's robustness may prove difficult because they are highly sensitive to small changes in user phrasing and to the language used. For instance, during summary generation, keywords such as "not" may be dropped, which can completely reverse the intended meaning of the advice.

Another reported limitation is the occurrence of errors under the form of completely erroneous generations that are often referred to as "perseveration" [13]. They occur when the required information is not available in the training dataset, and the model therefore fills in a plausible-sounding and coherently phrased fabrication. Unfortunately, these are indistinguishable from clinically valid information for a lay person. Overreliance on automated models should thus be avoided.

Finally, there is a lack of "privacy," as user-provided data are not protected under medical confidentiality regulations [12], and questions regarding liability and accountability are likely to arise.

## Conclusions and Future Perspectives

Patients increasingly consult AI chatbots for medical advice regarding their diagnosis and treatment. Phenomena observed in these interactions, including stereotyping, sycophancy, and perseveration, underscore the need for human oversight, the development of purpose-built chatbots for medical advice, and rigorous validation prior to deployment in clinical settings.

It is important to recognize, however, that some of these phenomena, such as sycophancy, are not unique to chatbot interactions and may equally occur in human communication. Moreover, in clinical practice, history-taking constitutes only 1 component of the diagnostic workup, complemented by findings from clinical neurological examination, medical imaging, and other investigations. Establishing their position within the existing acute care landscape is an essential first step, allowing future research to focus on meaningful comparisons.

Beyond their integration into existing care pathways, AI chatbots and their successors have the potential to transform medical practice as we know it. Caution must be taken, as blindly trusting them may lead to overlooked errors with important consequences, while overdiagnosis may lead to medical overconsumption. Given the ongoing need for human expertise and legal concerns regarding privacy and liability, we propose that health care professionals as well as local policymakers should be involved in the development and testing of LLMs' underlying AI chatbots. Furthermore, we should inform patients about risks and limitations.

Curiously but cautiously integrating these new tools into decision-making, information storage, and administrative support may pave the way for AI chatbots supporting clinicians. Assisting lay users to navigate this evolving landscape provides a novel challenge.

## Generative AI Use

ChatGPT 4.1 (OpenAI) and Grok 3 (xAI) were used to preserve spelling errors and nuances when translating the original chats.

## Funding

The authors declared no financial support was received for this work.



XSL•FO
RenderX

## Authors' Contributions

Conceptualization: EVD, FV
Methodology: EVD, FV
Investigation: SD, AM
Writing—original draft: SD, FV
Writing—review & editing: PS, EVD, HG, SMS, SDR, MVL

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Chat history for Case 1.
[DOCX File, 16 KB - i-jmr_v15i1e107318_app1.docx ]

Multimedia Appendix 2
Chat history for Case 2.
[DOCX File, 14 KB - i-jmr_v15i1e107318_app2.docx ]

Multimedia Appendix 3
Chat history for Case 3.
[DOCX File, 22 KB - i-jmr_v15i1e107318_app3.docx ]

Multimedia Appendix 4
MRI scan for Case 2.
[DOCX File, 150 KB - i-jmr_v15i1e107318_app4.docx ]

Checklist 1
CARE checklist.
[DOCX File, 18 KB - i-jmr_v15i1e107318_app5.docx ]

## References

1. Kothari RU, Pancioli A, Liu T, Brott T, Broderick J. Cincinnati Prehospital Stroke Scale: reproducibility and validity. Ann Emerg Med 1999 Apr;33(4):373-378. [doi: 10.1016/s0196-0644(99)70299-4] [Medline: 10092713]
2. Advani R, Naess H, Kurz M. Mass media intervention in Western Norway aimed at improving public recognition of stroke, emergency response, and acute treatment. J Stroke Cerebrovasc Dis 2016 Jun;25(6):1467-1472. [doi: 10.1016/j.jstrokecerebrovasdis.2016.02.026] [Medline: 27019989]
3. Darkhabani MZ, Homa-Bonell JK, Thoreson L, Bobholz JA, Spaulding D, Engebose M. BE FAST versus FAST: a randomized pilot trial comparing retention of stroke symptoms between 2 mnemonics. J Am Heart Assoc 2024 Oct;13(19):e035696. [doi: 10.1161/JAHA.123.035696] [Medline: 39291473]
4. ChatGPT. URL: https://chat.openai.com [accessed 2026-04-13]
5. SpaceXSI. URL: https://x.ai [accessed 2025-09-22]
6. de Boer JN, Voppel AE, Begemann MJ, Schnack HG, Wijnen F, Sommer IE. Clinical use of semantic space models in psychiatry and neurology: a systematic review and meta-analysis. Neurosci Biobehav Rev 2018 Oct;93:85-92. [doi: 10.1016/j.neubiorev.2018.06.008] [Medline: 29890179]
7. Ciampelli S, Voppel AE, de Boer JN, Koops S, Sommer IE. Combining automatic speech recognition with semantic natural language processing in schizophrenia. Psychiatry Res 2023 Jul;325:115252. [doi: 10.1016/j.psychres.2023.115252] [Medline: 37236098]
8. Luykx JJ, Gerritse F, Habets PC, Vinkers CH. The performance of ChatGPT in generating answers to clinical questions in psychiatry: a two-layer assessment. World Psychiatry 2023 Oct;22(3):479-480. [doi: 10.1002/wps.21145] [Medline: 37713576]
9. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 8;9:e45312. [doi: 10.2196/45312] [Medline: 36753318]
10. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023 Jun 1;183(6):589-596. [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]



XSL•FO
RenderX

11. Tu T, Schaekermann M, Palepu A, et al. Towards conversational diagnostic artificial intelligence. Nature 2025 Jun;642(8067):442-450. [doi: 10.1038/s41586-025-08866-7] [Medline: 40205050]
12. Fournier-Tombs E, McHardy J. A medical ethics framework for conversational artificial intelligence. J Med Internet Res 2023 Jul 26;25:e43068. [doi: 10.2196/43068] [Medline: 37224277]
13. Thirunavukarasu AJ, Ting DS, Elangovan K, Gutierrez L, Tan TF, Ting DS. Large language models in medicine. Nat Med 2023 Aug;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
14. Williams M, Carroll M, Narang A, Weisser C, Murphy B, Dragan A. On targeted manipulation and deception when optimizing LLMs for user feedback [Poster]. Presented at: International Conference on Learning Representations 2025; Apr 24-28, 2025 URL: https://openreview.net/pdf?id=Wf2ndb8nhf [accessed 2026-11-06]

## Abbreviations

**CT:** computed tomography
**FAST:** face drooping, arm weakness, speech difficulties, time to call emergency services
**LLM:** large language model
**MRI:** magnetic resonance imaging
**TIA:** transient ischemic attack

*Edited by TDA Cardoso; submitted 17.Jul.2026; peer-reviewed by M Toma, S Zhuang; comments to author 01.Sep.2026; revised version received 19.Oct.2026; accepted 23.Oct.2026; published 27.Nov.2026.*

*Please cite as:*
*Dierickx S, Speziali P, Gossye H, Slootjes SM, De Raedt S, Missiaen A, Van Lint M, Van Dellen E, Vandervorst F*
*Patients With Acute Stroke Symptoms Consulting Large Language Model Chatbots Before Presenting to the Emergency Department: Case Series*
*Interact J Med Res 2026;15:e107318*
*URL: https://www.i-jmr.org/2026/1/e107318*
*doi:10.2196/107318*

© Sara Dierickx, Paolo Speziali, Helena Gossye, Sofia Maldonado Slootjes, Sylvie De Raedt, Anaïs Missiaen, Michel Van Lint, Edwin Van Dellen, Fenne Vandervorst. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 27.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# An Eye-Opening Approach: Cancer of Unknown Primary Source With Choroidal Metastasis Case Report

Erin Kaser[1], DO; Vedin Barve[2], BS; Alexander Blaschke[3], DPM; Constance Fry[4], MD; Corey Waldman[5], MD; Sarah Hackman[6], MD; William Kelly[1], MD

[1]Department of Internal Medicine, The University of Texas Health Science Center at San Antonio, 7979 Wurzbach Road, San Antonio, TX, United States

[2]Long School of Medicine, The University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

[3]Division of Podiatry, Department of Orthopaedics, The University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

[4]Department of Opthalmology, University of Tennessee Health Science Center, Memphis, TN, United States

[5]Department of Opthalmology, The University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

[6]The University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

**Corresponding Author:**
William Kelly, MD
Department of Internal Medicine, The University of Texas Health Science Center at San Antonio, 7979 Wurzbach Road, San Antonio, TX, United States

## Abstract

Choroidal metastases (CM) represent a rare but clinically significant manifestation of systemic malignancy, most frequently from lung cancer. The choroid's vascular anatomy allows hematogenous tumor seeding. Although CM may be the first clinical sign of an underlying malignancy, evidence guiding its management in the modern immunotherapy era remains limited, as most published cases predate the widespread use of immune checkpoint-inhibitors. We describe a 33-year-old male patient presenting with ocular pain and visual disturbance, who was found to have an amelanotic choroidal lesion. Systemic workup revealed small pulmonary nodules and an iliac crest lesion. Sequential biopsies suggested that this was metastatic adenocarcinoma of unknown primary origin, but most likely of lung origin, without actionable mutations or PD-L1 (programmed death-ligand 1) expression. Management required multidisciplinary coordination and included carboplatin, paclitaxel, and pembrolizumab, followed by radiation to the orbit, iliac crest, and mediastinal sites of disease. Unfortunately, he experienced progression while on maintenance immunotherapy with new rib and brain lesions, for which he underwent treatment with platinum, pemetrexed, and bevacizumab with additional radiotherapy. Despite loss of vision in the affected eye, he achieved durable disease control and remains free of radiographic recurrent disease>4 years after diagnosis. This case illustrates that multimodality salvage strategies—integrating systemic therapy with aggressive local radiation—can provide unexpectedly prolonged survival even after immunotherapy failure. Importantly, current guidelines offer minimal direction on managing CM in this context, and prior case reports do not reflect present-day treatment realities. The key message for clinicians is that CM should not automatically be approached with palliative intent; carefully selected patients may benefit from an oligometastatic strategy that actively targets limited metastatic sites to prolong survival. Our findings underscore the need for ophthalmology, radiation oncology, and medical oncology collaboration when vision-threatening or occult metastatic lesions arise. For readers, the takeaway is that choroidal metastasis—particularly in the era of immunotherapy—warrants individualized, multidisciplinary evaluation rather than default palliation. Our case demonstrates that coordinated multimodality management can achieve long-term disease control, highlighting a treatment paradigm worth considering for selected patients and calling for updated guidelines that reflect modern therapeutic capabilities.

***(Interact J Med Res 2026;15:e77895)*** doi:10.2196/77895

**KEYWORDS**

choroidal; metastasis; non-small cell lung cancer; cancer of unknown primary origin; multi-modal therapy

## Introduction

Rare presentations of cancer can include metastases to the orbit through hematogenous dissemination. The uveal tract is a highly vascular complex made from the choroid, ciliary body, and iris. Importantly, the eye contains the blood-retinal barrier, analogous to the blood-brain barrier. However, the choroid layer lies outside retinal pigment epithelial tight junctions and is therefore exposed to unimpeded drug delivery from the vasculature [1]. One of the most common sources of choroidal metastases is primary lung cancer, with 88% of lung metastases to the eye involving the choroid, with an associated overall survival at one-year of 54% [2]. Consequently, choroidal metastases (CM) of unknown origin are often treated similarly to lung cancer as



XSL•FO
RenderX

the clinical, imaging, and histopathologic features of choroidal lesions frequently resemble those seen in lung cancer patients [3,4]. The prognosis for carcinoma of unknown primary (CUP) with CM is generally poor, with median overall survival ranging from 3 to 12 months, as patients tend to present with advanced and disseminated disease [5,6].

## Case Presentation

### Patient Information

The patient was a 33-year-old male at the time of initial presentation with no significant past medical history. He presented with right ophthalmalgia, blurry vision, and sinus congestion. Clinical course was as follows (Table 1).

**Table .** Timeline of key events in the described patient presentation

| Day | Description of key events. |
|---|---|
| Day 0-2 | Initial presentation: chief complaint of right ophthalmalgia and blurry vision. Ophthalmology observed an amelanotic right choroidal lesion. MRI[a] Orbit showed enhancing plaque of the posterior-lateral wall of right globe. CT[b] chest showed RML and RUL nodules |
| Day 44-113 | Pathological workup: right intraophthalmic choroidal biopsy with scant atypical cells. PET (Positron Emission Tomography) revealed non-avid RML[c] and RUL[d] nodules and mildly avid L iliac crest. R choroidal FNA[e] read as metastatic carcinoma, compatible with pulmonary adenocarcinoma. L iliac biopsy with further suggestion of lung origin (CK7, TTF1+) |
| Day 125-313 | Initial chemoimmunotherapy: started on carboplatin, paclitaxel and pembrolizumab for 4 cycles followed by pembrolizumab maintenance. Completed stereotactic radiation to R orbit |
| Day 293-324 | Disease progression: clinical and radiographic progression at ribs and lungs as well as intracranial site |
| Day 313-743 | Oligometastatic salvage therapy: given radiation to chest wall and left frontal operculum. Started on carboplatin, pemetrexed, and bevacizumab for 4 cycles followed by pemetrexed and bevacizumab maintenance. Bevacizumab was continued as maintenance for 14 cycles |
| Day 743-Present | Surveillance: no clinical or radiographic evidence of disease progression as of day 1563. |

[a]MRI: magnetic resonance imaging.

[b]CT: computed tomography.

[c]RML: right middle lobe.

[d]RUL: right upper lobe.

[e]FNA: fine-needle aspiration.

## Ethical Considerations

The patient involved in this case report has given his informed consent authorizing the use and disclosure of his health information.

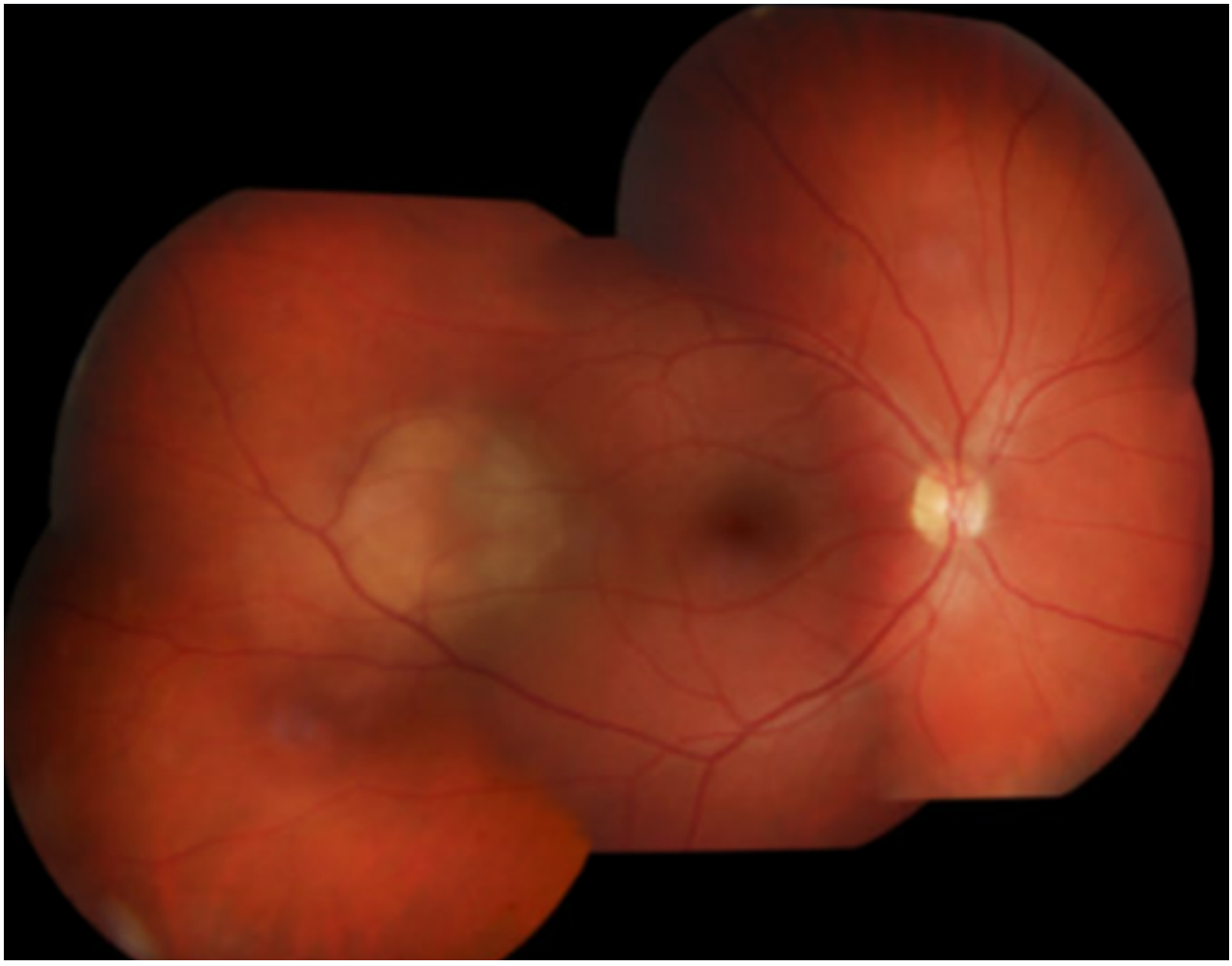
## Clinical Findings

Evaluation by ophthalmology revealed an amelanotic right choroidal lesion (1.04 × 1.15×.31 cm, Figure 1). The patient had 20/20 - 1 vision bilaterally with full visual fields.



XSL•FO
RenderX

**Figure 1.** Ophthalmologic evaluation showed an amelanotic right choroidal lesion.



## Diagnostic Assessment

MRI Orbit on Day 0 (Figure 2A) confirmed an enhancing plaque along the posterior-lateral wall of his right globe. A CT Chest (Day 2) demonstrated a right middle lobe (RML) (0.5cm) and right upper lobe (RUL) nodule (0.2cm). The patient underwent intraophthalmic choroidal biopsy (Day 44). Pathology read as scant atypical cells, with few cells showing 'neuroendocrine-like nuclear features.' A DOTATATE PET (Positron Emission Tomography) on Day 64 was performed, which showed mild avidity in the right globe mass (SUV 2.6, Krenning 1) (Figure 3A), a nonavid RML nodule (0.5cm, below PET resolution) and RUL nodule (0.2cm), and a mildly avid left iliac crest lesion (1cm) (Figure 3B, C). The patient subsequently underwent a right choroidal fine-needle aspiration (FNA) on Day 69, with pathology read as metastatic carcinoma, compatible with pulmonary adenocarcinoma. Notably, the histology did not resemble a retinal pigment epithelium tumor and was positive for CK7 and TTF1 but negative for PD-L1, thyroglobulin, PAX8, CD20, GATA3, CD3, and Melan-A. CK20 was only weakly focally positive. Synaptophysin highlighted pieces of retina but not the tumor. Molecular testing showed no actionable mutation in EGFR, KRAS, BRAF, or HER2. Unfortunately, testing for ALK, RET, ROS1, NTRK1/3 and MET was inconclusive due to insufficient sample. A left iliac biopsy (Day 113, Figure 4) and mediastinal lymph node (4R) FNA biopsy (Day 114) were also read as carcinoma showing CK7 and TTF1 positivity. An insufficient sample of tumor was procured for sequencing; however, immunohistochemical (IHC) analysis was negative for EGFR, ALK, ROS, and BRAF driver mutations. No PD-L1 expression (0%) was observed. The patient was discussed in a multidisciplinary tumor board with the ultimate determination that this was metastatic adenocarcinoma of unknown primary, but most likely of pulmonary origin and to treat as non-small cell lung cancer (NSCLC).



XSL•FO
RenderX

**Figure 2.** (A) MRI orbit (Day 0) showed enhancing plaque along the posterior-lateral wall of his right globe. (B) MRI orbit (Day 1,123) showed new right optic nerve enhancement.

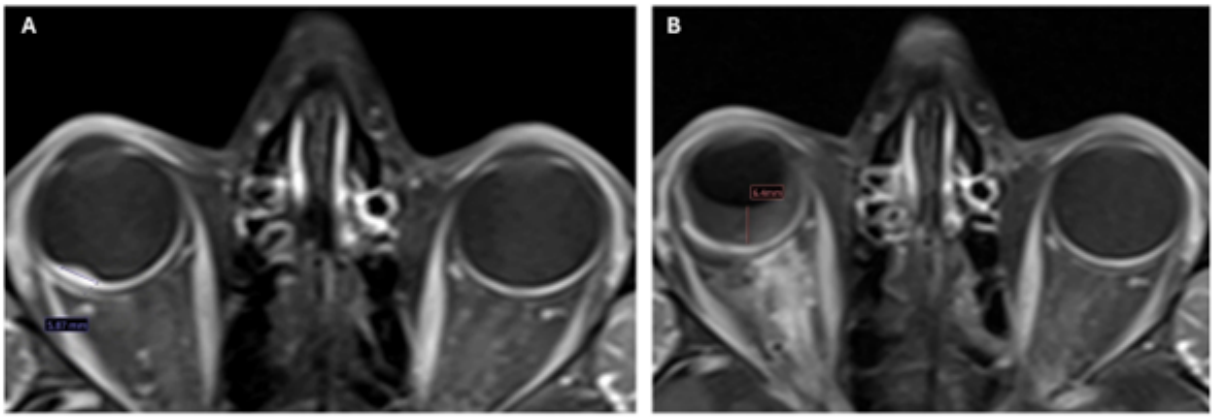
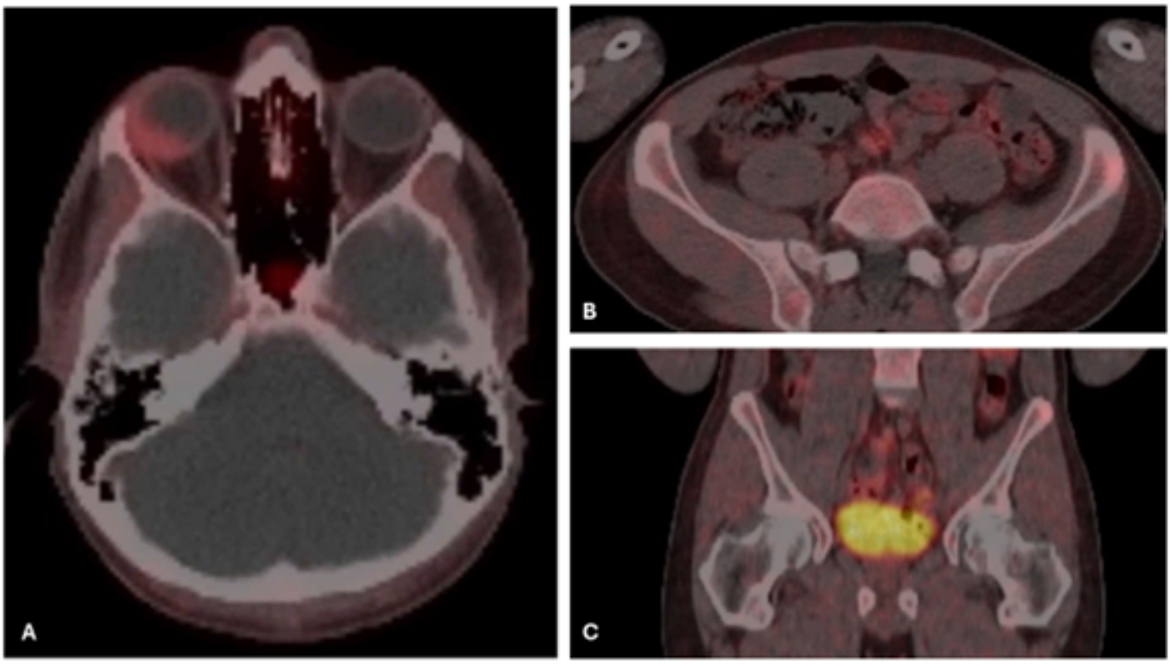


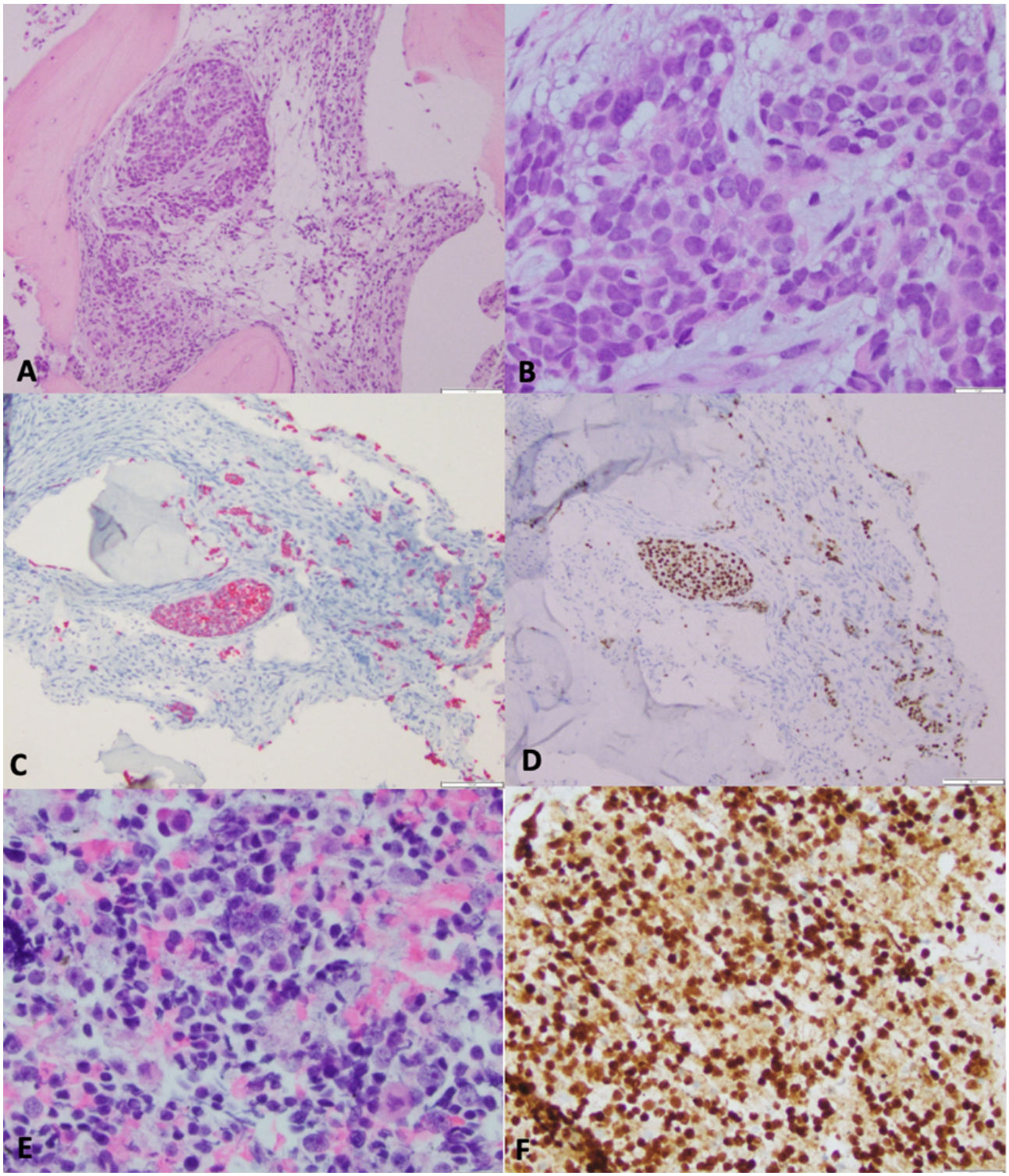

**Figure 3.** DOTATATE PET Imaging on d 64 demonstrating sites of metastatic involvement(A) Axial DOTATATE PET of the orbits showing mild radiotracer avidity within the right choroidal lesion (SUV 2.6; Krenning score 1), corresponding to the enhancing plaque previously identified on MRI; (B) Axial PET of the pelvis showing a mildly avid lytic lesion in the left iliac crest (approximately 1 cm), consistent with metastatic involvement; (C) Coronal PET confirming the left iliac crest lesion with focal uptake.






XSL•FO
RenderX

**Figure 4.** Iliac crest biopsy (Day 113) and pulmonary lymph node biopsy (Day 114). (A, B) Iliac crest bone lesion, H&E, 10 x and 40 x: tumor cells with bone; (C) Iliac crest bone lesion, Cytokeratin AE1/3 immunostaining, 10 x: cytokeratin stain confirming the tumor cells are carcinoma; (D) Iliac crest bone lesion, TTF-1 immunostaining, 10 x: TTF-1 stain supports lung origin; (E) Pulmonary lymph node: H&E, 40 x; tumor cells are morphologically comptabile with those in bone; (F) pulmonary lymph node: TTF-1 immunostaining, 20 x: TTF-1 stain supports lung origin. TTF: thyroid transcription factor.




## Therapeutic Intervention

Following tumor board discussion, treatment was initiated with carboplatin (AUC 6), paclitaxel (200mg/m$^2$), and pembrolizumab (200mg) every 21 days (C1D1 on Day 125) for 4 cycles, with pembrolizumab continued as maintenance after cycle 4. Alternative therapies considered included a palliative approach with chemotherapy alone, but age, functional status, limited extent of disease, and the patient's wishes to pursue an aggressive therapeutic approach led to the chosen treatment protocol. Paclitaxel was chosen over pemetrexed as the taxane used because it was deemed to cover more histologies in the event that the cancer was not actually of bronchopulmonary origin. Patient completed stereotactic radiosurgery (SRS) (Day



XSL•FO
RenderX

173) with 36 Gray (Gy) in 4 fractions (fx) to the right orbit lesion. Following cycle 4, he was placed on pembrolizumab maintenance. PET (Day 211) showed an interval increase in L iliac crest lesion and increased conspicuity of R lung base grouped nodularity, but otherwise stable R paratracheal (0.7cm) and mediastinal (0.4cm) lymph nodes and a decreased RML (0.3cm) pulmonary nodule. He also completed radiation to the left iliac (Day 252, 50Gy in 5 fx), RML (Day 252, 50Gy in 5 fx), and mediastinal nodes (Day 273, 40Gy in 10 fx). However, he began having rib pain and PET/CT (Day 293) showed a new lesion in the left 4th rib and RLL, as well as an increase in additional right pulmonary nodules, concerning for progression. Concurrently, MRI Brain (Day 324) had also shown a new left frontal gyrus lesion. Given his age and continued limited disease involvement, an oligometastatic approach was used, and the patient completed radiation to the left chest wall (Day 313, 50Gy in 5 fx) and left frontal operculum (Day 337, 1 fx). Pembrolizumab was stopped (last dose on Day 315) due to radiographic progression, and he was started (C1D1 on Day 344) on carboplatin (AUC 5), pemetrexed (500mg/m$^2$), and bevacizumab (15mg/kg) every 21 days for 4 cycles, with pemetrexed and bevacizumab continued as maintenance after cycle 4. Pemetrexed and bevacizumab were held after 14 cycles (last dose on Day 743) in preparation for cataract removal of the right eye, and he was subsequently placed on surveillance with MRI Brain and PET imaging every 3 months.

## Follow-Up and Outcomes

He developed eye pain, and MRI orbit (Day 1123, Figure 2B) showed right optic nerve enhancement, which was new from previous imaging and felt secondary to prior radiotherapy. He received one dose of intraocular bevacizumab injection, which completely resolved the pain. The patient had developed blindness in his right eye after initial treatment, but remained with 20/20 vision on the left. He maintained his employment and continues with regular follow-up and surveillance imaging every 3 months. His scans as of Day 1563 show no evidence of recurrence.

## Discussion

Given CUP with CM tends to present with late, advanced, and disseminated disease, the general treatment protocol is palliative, focusing on symptom control and vision preservation. For most patients, external beam radiotherapy is the standard local treatment for symptomatic choroidal metastases, with systemic therapy including platinum- or taxane-based chemotherapy used for widespread disease [7]. Systemic therapy for CM of CUP often mirrors lung cancer regimens because empirical chemotherapy and targeted therapies have demonstrated efficacy in controlling both systemic and ocular disease in lung cancer patients [8-10]. For metastatic non-small cell lung cancer, platinum-based chemotherapy with PD-1 blockade is the standard first-line treatment; however, data regarding CM in the age of immunotherapy are limited. Recently, Li et al presented a case of a 65-year-old male patient with NSCLC and CM who demonstrated ultrasonic regression, improved visual acuity, and symptomatic relief after treatment with pemetrexed and capecitabine plus pembrolizumab [11]. A second case report by Matsuyama et al describes treatment with nivolumab plus ipilimumab plus chemotherapy, which resulted in symptomatic improvement and no relapse at 18 months follow-up [12]. These isolated cases underscore both the rarity of CM and the need for evidence-guided management strategies.

Bevacizumab, an extracellular VEGF-A inhibitor, has strong anti-angiogenic properties and has been used to treat CM from pulmonary, breast, and colorectal metastases. For instance, Riess JW et al described a case series of 3 patients with NSCLC and CM who were treated with a backbone of platinum-based chemotherapy plus bevacizumab [13]. The first patient received SRS to the symptomatic CM with improved vision and 12 months of progression-free survival. The second patient was found to have CM after right eye vision loss. This vision normalized following whole-brain radiotherapy that included the choroidal regions. Thereafter, the patient received 14 cycles of chemotherapy followed by erlotinib for a discovered EGFR mutation and progression. The third patient tolerated 16 months of chemotherapy followed by a recurrence of CM, which was treated with an ALK inhibitor for a found translocation. Furthermore, George et al reported on a 42-year-old female patient diagnosed with NSCLC with CM who achieved complete response of her 1.2cm CM following 3 cycles of carboplatin, gemcitabine, and bevacizumab, and was still alive with reported normal vision following her 7th maintenance dose of bevacizumab [14].

In the case of our patient, timely referral to ophthalmology led to recognition of a uveal lesion within three weeks of symptom onset. PET imaging can be helpful in defining dominant lesions and disease burden but histologic evaluation is necessary to guide diagnosis. In situations such as ours, where the primary site cannot be readily identified from imaging, judicious use of immunohistochemistry and molecular profiling is called for. In the era of immunotherapeutics, the median survival of newly diagnosed NSCLC without targetable mutation ranges from 14 to 22 months [15,16]. At the time of writing, our patient has survived 4 years with no radiographic or clinical signs of recurrence. While we must be careful not to read too much into an isolated case, it is the authors' belief that an aggressive oligometastatic approach with multimodality salvage options may be best for all NSCLC patients or patients with suspected primary lung malignancy who initially present with choroid disease and limited disease elsewhere.

As presented, this case illustrates an innovative management strategy for choroidal metastasis in the era of immunotherapy. While prior reports describe short-term responses to chemotherapy, anti-VEGF therapy, or immune checkpoint inhibitors, none have documented long-term survival following an oligometastatic approach that was applied both at diagnosis and after systemic progression. Our patient's survival of more than 4 years exceeds historical expectations for metastatic NSCLC without targetable alterations or PD-1 expression and highlights the potential benefits of aggressive local therapy even at the cost of vision.

The substantial collaboration required to execute this treatment approach is equally important. Coordination between imaging,



XSL•FO
RenderX

multiple ophthalmologic biopsies, and subspecialty pathologic review was essential for diagnosis. Treatment required communication between medical and radiation oncology to prioritize systemic control while delivering targeted radiation to multiple metastatic sites. This approach reflects the real-world practice and serves as a model for clinicians managing similarly rare and complex presentations.

## Conclusion

CM remains an uncommon presentation of systemic malignancy, and contemporary management is challenged by the absence of evidence-based guidelines that reflect the current era of immunotherapy. Our case demonstrates that even after progression on checkpoint inhibitors, carefully selected patients may achieve prolonged survival through a coordinated multimodality approach that integrates systemic therapy with site-directed radiation. This experience suggests that CM should not automatically be considered a purely palliative entity, particularly when disease burden is low and lesions are anatomically accessible for local therapy.

More broadly, this report emphasizes the importance of early ophthalmologic evaluation for unexplained visual symptoms, the value of thorough systemic staging when CM is suspected, and the critical role of multidisciplinary collaboration involving medical oncology, radiation oncology, and ophthalmology. As prior published cases predate modern immunotherapy, this case highlights an unmet need for updated clinical guidance on managing CM in patients who have progressed on or failed to respond to immune checkpoint inhibitors.

### Acknowledgments

No Generative AI was used throughout the generation of this manuscript.

### Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

### Authors' Contributions

EK and WK were involved in the preparation and writing of the initial draft. VB reviewed and edited the initial draft and prepared the manuscript for submission. WK provided overall supervision of the development of the manuscript. AB, CF, CW, and SH were involved in providing histology and biopsy images as well as reviewing the original manuscript for revisions.

### Conflicts of Interest

None declared.

### References

1. Yang X, Yu XW, Zhang DD, Fan ZG. Blood-retinal barrier as a converging pivot in understanding the initiation and development of retinal diseases. Chin Med J (Engl) 2020 Nov 5;133(21):2586-2594. [doi: 10.1097/CM9.0000000000001015] [Medline: 32852382]
2. Kim RS, Nguyen MK, Card KR, Shields CL. Utility of systemic brigatinib therapy in tumour recurrence of choroid metastasis from non-small cell lung carcinoma. Can J Ophthalmol 2024 Aug;59(4):e397-e399. [doi: 10.1016/j.jcjo.2023.04.013] [Medline: 37236265]
3. Shields CL, Kalafatis NE, Gad M, et al. Metastatic tumours to the eye. Review of metastasis to the iris, ciliary body, choroid, retina, optic disc, vitreous, and/or lens capsule. Eye (Lond) 2023 Apr;37(5):809-814. [doi: 10.1038/s41433-022-02015-4]
4. Blasi MA, Maceroni M, Caputo CG, et al. Clinical and ultrasonographic features of choroidal metastases based on primary cancer site: Long-term experience in a single center. PLoS One 2021;16(3):e0249210. [doi: 10.1371/journal.pone.0249210] [Medline: 33765045]
5. Jardel P, Sauerwein W, Olivier T, et al. Management of choroidal metastases. Cancer Treat Rev 2014 Dec;40(10):1119-1128. [doi: 10.1016/j.ctrv.2014.09.006] [Medline: 25451606]
6. Mathis T, Jardel P, Loria O, et al. New concepts in the diagnosis and management of choroidal metastases. Prog Retin Eye Res 2019 Jan;68:144-176. [doi: 10.1016/j.preteyeres.2018.09.003] [Medline: 30240895]
7. Raghav K. Cancer of unknown primary site. N Engl J Med 2025 May 29;392(20):2035-2047. [doi: 10.1056/NEJMcp2402691] [Medline: 40435465]
8. Yang CJ, Tsai YM, Tsai MJ, Chang HL, Huang MS. The effect of chemotherapy with cisplatin and pemetrexed for choroidal metastasis of non-squamous cell carcinoma. Cancer Chemother Pharmacol 2014 Jan;73(1):199-205. [doi: 10.1007/s00280-013-2341-4] [Medline: 24202667]
9. Chee ASH, Mak ACY, Kam KW, et al. Efficacy of epidermal growth factor receptor tyrosine kinase inhibitor in treating choroidal metastasis from non-small-cell lung cancer: a 10-year review. Retina 2025 Feb 1;45(2):207-214. [doi: 10.1097/IAE.0000000000004314] [Medline: 39504451]



XSL•FO
RenderX

10. Maller B, Salvatori S, Tanvetyanon T. Outcomes of intraocular metastasis from lung cancer in the era of targeted therapy: a systematic review and pooled analysis. Clin Lung Cancer 2022 Dec;23(8):e519-e525. [doi: 10.1016/j.cllc.2022.07.018] [Medline: 36030188]
11. Li X, Liang Y, Wang J, Hu Z, Yuan ZL, Xie P. Pembrolizumab for choroidal metastases from non-small cell lung cancer: a case report and literature review. Hum Vaccin Immunother 2021 Aug 3;17(8):2626-2629. [doi: 10.1080/21645515.2021.1885282] [Medline: 33769211]
12. Matsuyama T, Oniwa M, Tsuruzono K, et al. Improving visual acuity with nivolumab plus ipilimumab plus two cycles of chemotherapy following a diagnosis of lung adenocarcinoma with choroidal metastasis: a case report and literature review. Respirol Case Rep 2024 Jan;12(1):e01262. [doi: 10.1002/rcr2.1262] [Medline: 38045825]
13. Riess JW, Nagpal S, Das M, Neal JW, Kim JW, Wakelee HA. A case series of NSCLC patients with different molecular characteristics and choroidal metastases: improvement in vision with treatment including pemetrexed and bevacizumab. J Thorac Oncol 2013 Feb;8(2):e17-e18. [doi: 10.1097/JTO.0b013e31827690da] [Medline: 23328555]
14. George B, Wirostko WJ, Connor TB, Choong NW. Complete and durable response of choroid metastasis from non-small cell lung cancer with systemic bevacizumab and chemotherapy. J Thorac Oncol 2009 May;4(5):661-662. [doi: 10.1097/JTO.0b013e31819c9a73] [Medline: 19395911]
15. Garassino MC, Gadgeel S, Speranza G, et al. Pembrolizumab plus pemetrexed and platinum in nonsquamous non-small-cell lung cancer: 5-year outcomes from the phase 3 KEYNOTE-189 study. J Clin Oncol 2023 Apr 10;41(11):1992-1998. [doi: 10.1200/JCO.22.01989] [Medline: 36809080]
16. Johnson ML, Cho BC, Luft A, et al. Durvalumab with or without tremelimumab in combination with chemotherapy as first-line therapy for metastatic non-small-cell lung cancer: the phase III POSEIDON study. J Clin Oncol 2023 Feb 20;41(6):1213-1227. [doi: 10.1200/JCO.22.00975] [Medline: 36327426]

## Abbreviations

**ALK:** anaplastic lymphoma kinase
**AUC:** area under the curve
**BRAF:** B-rapidly accelerated fibrosarcoma (BRAF) gene
**CK7:** Cytokeratin 7
**CM:** choroidal metastases
**CT:** computed tomography
**CUP:** cancer of unknown primary
**CxDy:** Cycle x Day y
**DOTATATE PET:** DOTA-Tyr³-octreotate Positron Emission Tomography
**EGFR:** epidermal growth factor receptor
**FNA:** fine-needle aspiration
**Fx:** fraction
**Gy:** gray (unit of radiation dose)
**IHC:** immunohistochemistry
**MRI:** magnetic resonance imaging
**NSCLC:** non-small cell lung cancer
**PAX8:** paired box gene 8
**PD-L1:** programmed death-ligand 1
**RLL:** right lower lobe
**RML:** right middle lobe
**ROS1:** ROS proto-oncogene 1 receptor tyrosine kinase
**RUL:** right upper lobe
**SRS:** stereotactic radiation
**TTF1:** thyroid transcription factor-1
**VEGF-A:** vascular endothelial growth factor – A

*Edited by A Schwartz, M Balcarras; submitted 22.May.2025; peer-reviewed by T Matsuyama; accepted 02.Jan.2026; published 05.Feb.2026.*

*Please cite as:*
*Kaser E, Barve V, Blaschke A, Fry C, Waldman C, Hackman S, Kelly W*
*An Eye-Opening Approach: Cancer of Unknown Primary Source With Choroidal Metastasis Case Report*
*Interact J Med Res 2026;15:e77895*
*URL: https://www.i-jmr.org/2026/1/e77895*
*doi:10.2196/77895*



XSL•FO
RenderX

© Erin Kaser, Vedin Barve, Alexander Blaschke, Constance Fry, Corey Waldman, Sarah Hackman, William Kelly. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 5.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Case Report

# Orthodontic Management of Vertically Impacted Mandibular Permanent Second Molar in Adults: Case Report

Mashaka Hassan Mrisho[1,2*], MDent; Cui Lingling[1*], Prof Dr; Mary Paul Pallangyo[3*], MD; Jiaao Li[1*], MDent; Deogratius David Galikano[4*], MD; Xinwei Guo[1*], MDent; Yue Zhang[1*], MDent; Si Long Xiao[1*], MDent; Yan Zhang[1*], DDS

[1]Department of Orthodontics, First Affiliated Hospital, Jinzhou Medical University, Jinzhou City, Liaoning, China

[2]Department of Orthodontics Paedodontics and Community Dentistry, School of Dentistry, Muhimbili University of Health and Allied Sciences, Dar Es Salaam, United Republic of Tanzania

[3]Medical Department, Southern Zone Referral Hospital, Mtwara, United Republic of Tanzania

[4]Medical Department, Mzumbe University, Morogoro, United Republic of Tanzania

[*]all authors contributed equally

**Corresponding Author:**
Cui Lingling, Prof Dr
Department of Orthodontics
First Affiliated Hospital
Jinzhou Medical University
Songpo Road Section 3
Jinzhou City, Liaoning, 121000
China
Phone: 86 15941608206
Email: cuiawenn@163.com

## Abstract

Impacted mandibular permanent second molars (MM2) are uncommon, with a reported prevalence of 1.8% to 3% among orthodontic populations, and vertically impacted MM2 are particularly infrequently reported. This case describes the orthodontic preservation of a vertically impacted tooth 47 in a 20-year-old Chinese male who presented with recurrent right-sided mandibular pain, gingival swelling, and bleeding associated with pericoronitis around tooth 48. Clinical examination revealed an Angle Class I molar relationship on the left and a Class II relationship on the right, a rightward mandibular midline deviation, 3 mm of anterior mandibular crowding, overeruption of teeth 17 and 27, and an unerupted tooth 47. Panoramic imaging showed vertically impacted teeth 18, 28, and 47, a horizontally impacted tooth 48, and mesioangular impaction of tooth 38. Cone beam computed tomography demonstrated the relationship between tooth 47 and adjacent structures and minimal resorption of the distal root of tooth 46, without obvious root dilaceration or clear ankylosis-related changes. Cephalometric findings included a skeletal Class I relationship, bimaxillary protrusion, proclined incisors, and a hypodivergent facial pattern. Following a discussion of extraction and implant replacement versus orthodontic retrieval, the patient selected preservation for psychosocial and economic reasons, with extraction planned if no movement occurred within 2 to 3 months. Treatment included fixed appliances, extraction of tooth 48, and surgical exposure of tooth 47. At month 6, with 0.017×0.025-inch stainless steel archwires in place, a customized 0.7-mm stainless steel loop with an approximately 10-mm distal extension was bonded to tooth 46. Elastic traction between this loop and the attachment on tooth 47 applied approximately 50 to 60 gf of extrusive force. Tooth 47 became visible at month 8, approximately half its crown was visible at month 11, and a buccal tube was then bonded for continued alignment. The tooth reached the occlusal level at month 13, and treatment was completed at month 14. Teeth 46 and 47 retained vital pulp responses, and no symptoms or apical pathologies were reported. Immediately after treatment, external root resorption of approximately 1 mm was observed on tooth 46; at 10 months, the corresponding radiolucent area was approximately 0.5 mm, with stable clinical findings and good oral hygiene. The outcome suggests that surgical exposure combined with tooth-borne orthodontic extrusion may be a feasible tooth-preserving option for carefully selected adults with vertically impacted mandibular second molars. Longer-term reports with standardized clinical and radiographic assessments are needed to define prognosis and broader applicability.

***(Interact J Med Res 2026;15:e101851)*** doi:10.2196/101851

**KEYWORDS**

mandible; tooth; impacted molar; orthodontic extrusion; case report



XSL•FO
RenderX

## Introduction

Impacted mandibular second molars (MM2s) are an uncommon clinical finding, with reported global incidence rates ranging from 0.03% to 1.7% [1-3] and slightly higher prevalence (1.8%-3%) among orthodontic populations [4]. These impactions frequently present diagnostic challenges as they are often asymptomatic and incidentally detected during routine radiographic examination. The primary etiological factors include insufficient retromolar space and local obstructive anomalies, specifically posterior crowding, reduced mandibular arch length, premature loss of the adjacent first molar, early eruption of the third molar, abnormal inclination of the second molar bud, and ankylosis [4-7]. Additionally, class II skeletal relationships associated with micrognathia and a reduced gonial angle have been implicated in MM2 impaction development [6]. Epidemiological studies indicate a male predilection, a peak detection age between 11 and 14 years, and a higher prevalence among individuals of Chinese ethnicity (2- to 3-fold greater than in other racial groups), with unilateral involvement more common than bilateral presentation [1,4].

MM2 impactions are classified into 3 distinct angulation patterns. Shpack et al [8], in their analysis of 165 impacted MM2s, reported that the majority (88%) were mesially angulated, with 8% being vertical and 4% distally angulated. This distribution underscores the relative rarity of vertically impacted MM2s, which have received limited attention in the existing literature.

Multiple treatment strategies have been described for managing impacted MM2s, including surgical exposure combined with orthodontic uprighting, surgical uprighting alone, extraction with subsequent mesial drift of the third molar, and extraction with autotransplantation of the maxillary or mandibular third molar (MM3) [9,10]. Treatment success depends upon several prognostic factors, including patient age, tooth inclination, root development stage, caries status, periodontal condition, root resorption of adjacent teeth, intraoral visibility, and surrounding bone support. Patients around 14 years of age demonstrate significantly more favorable outcomes, and mesially angulated impactions generally respond better to treatment than other types [2,9,11].

However, vertically impacted MM2s represent a distinct clinical entity that challenges conventional understanding of impaction etiology and management. Unlike mesially or distally angulated impactions, vertical impactions cannot be attributed solely to differential root growth patterns or space deficiency. Furthermore, existing treatment protocols and mechanical designs are predominantly oriented toward uprighting angulated impactions, leaving a notable gap in evidence-based guidance for vertical presentations.

This case report describes the successful orthodontic management of a vertically impacted MM2 in a 20-year-old Chinese male patient with a simple tooth-borne orthodontic appliance. The purpose of this case report is to provide a possible mechanical approach that can be used in the management of vertically impacted permanent MM2s in adults.

## Ethical Considerations

This case report was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from the patient before treatment. The patient was provided with signed consent for publication.

## Patient Information

A 20-year-old Chinese male patient presented to the Orthodontic Department of the First Affiliated Hospital of Jinzhou Medical University with a chief concern of intermittent mild pain on the right side of the lower jaw for about a week, associated with swollen gum around the last tooth on the same side and occasional bleeding during toothbrushing. No medication was taken for pain management. He reported that he had been experiencing intermittent swelling and pain on that side for nearly 6 months. The pain sometimes made chewing uncomfortable on the affected side. He denied a history of orofacial injury or previous orthodontic treatment. There was no history of chronic illnesses or a similar dental anomaly in the family. He was the only child in his family and was a high school student.

The patient was generally healthy with a normal gait and posture. Extraoral examination revealed a symmetrical straight facial profile with competent lips. The temporomandibular joint was normal, with no pain or clicking sound on palpation and no jaw deviation on opening or closing the mouth. No palpable swelling was detected on the affected side. Intraoral examination showed permanent dentition, an Angle class I molar relationship on the left and a class II relationship on the right side, a mandibular midline shift to the right, overeruption of teeth 17 and 27, an unerupted tooth 47 with no bulginess of the overlying mucosa, and pericoronitis on tooth 48 (Figure 1).

Model analysis showed an Angle class II molar relationship on the right side with a class I relationship on the left, 3 mm of crowding of the anterior mandibular teeth, lingually inclined teeth 37 and 42, and normal overjet and overbite (Figure 2).

Panoramic radiography showed vertical impaction in teeth 18, 28, and 47, a horizontally impacted tooth 48, mesial angulated impaction of tooth 38 and overeruption of teeth 17 and 27. The distal aspect of the MM2 crown was blocked by one third of MM3 mesial occlusal surface and its mesial aspect was in contact with the distal surface of the mandibular first molar (MM1) distal root with signs of minimal external root resorption at the contact point (Figure 3A and 3B) and no contact further down the root (Figure 3C). No signs of ankylosis areas were evident, and no root dilaceration were detected.

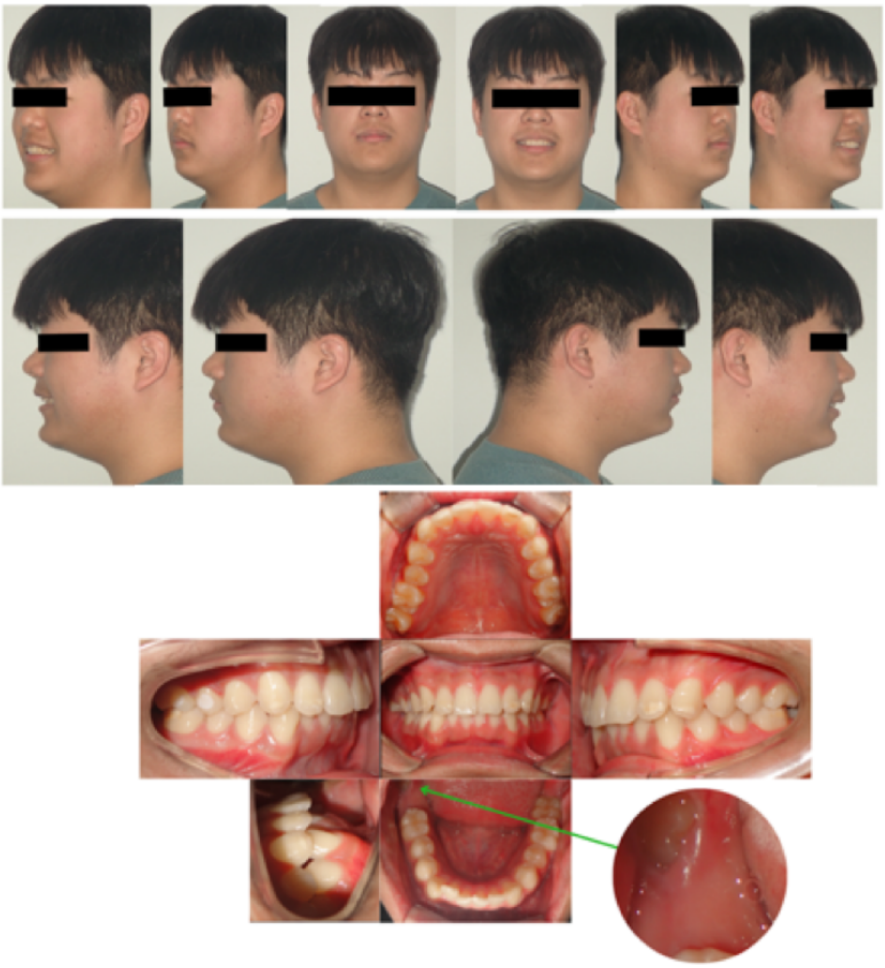
Cephalometric analysis according to Chinese norms showed increased sella turcica-nasion-point A (SNA) and sella turcica-nasion-point B (SNB) angles, normal point A-nasion-point B (ANB) angle, a decreased upper incisor–lower incisor (Ui-Li) angle, increased lower incisor–mandibular plane (Li-MP) angle, increased upper incisor-sella turcica-nasion (Ui-SN) angle, and a decreased mandibular plane-sella turcica-nasion (MP-SN) angle (Figure 4).



XSL•FO
RenderX

A diagnosis of Angle class II subdivision malocclusion with skeletal class I; vertically impacted teeth 18, 28, and 47; resorption of tooth 46 distal root surface; pericoronitis on a horizontally impacted tooth 48; mild crowding; overeruption of teeth 17 and 27; lingually inclined teeth 37 and 42; bimaxillary protrusion; proclination of both upper and lower incisors; and a hypodivergent face was made.

**Figure 1.** Extraoral photos (top) showing a symmetric face with a straight facial profile and competent lips. Intraoral photos (bottom) showing an unerupted tooth 47, overeruption of teeth 17 and 27, with mild crowding of the mandibular anterior teeth, and pericoronitis on tooth 48.





XSL•FO
RenderX

**Figure 2.** Study model showing an Angle class II molar relationship on the right side (red line), lingually inclined tooth 37, and crowding on the mandibular anterior teeth (measured 3 mm) with subsequent lingual inclination of tooth 42.

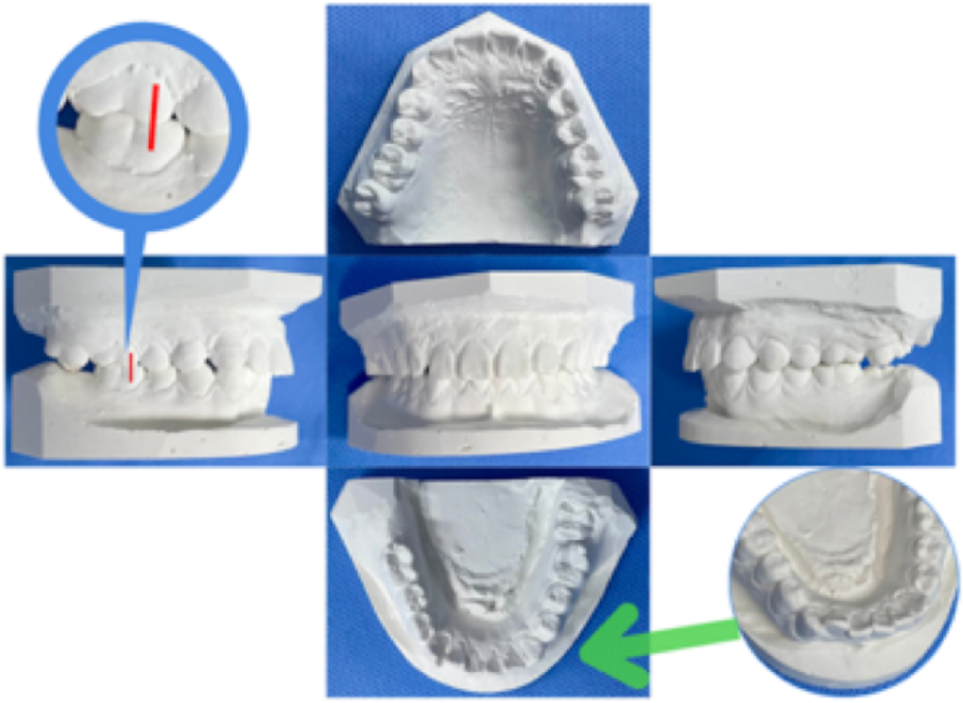
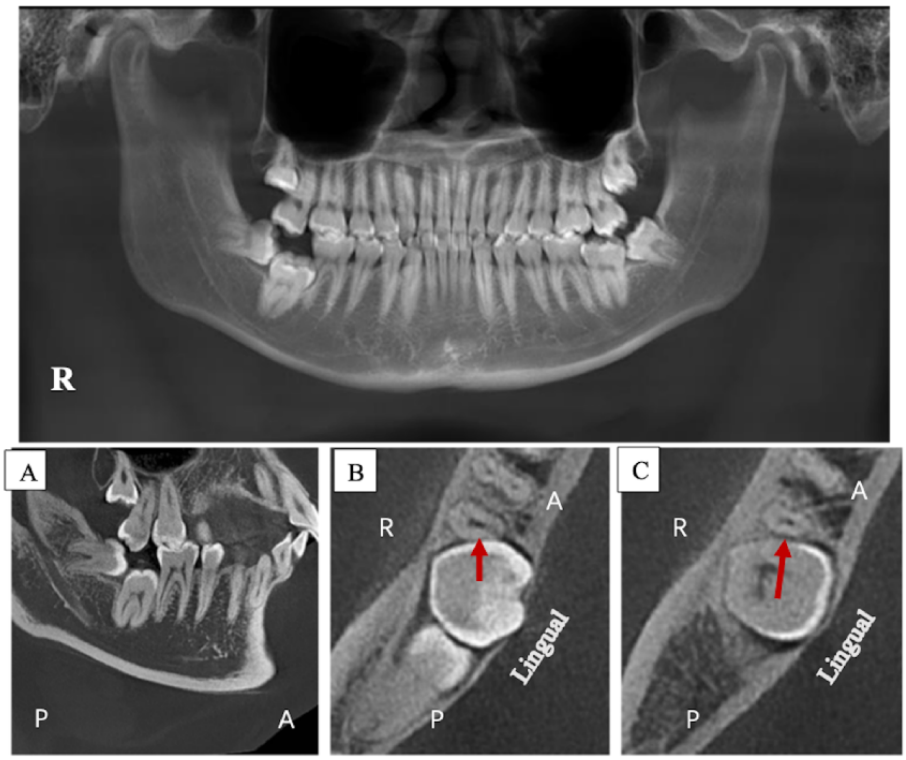


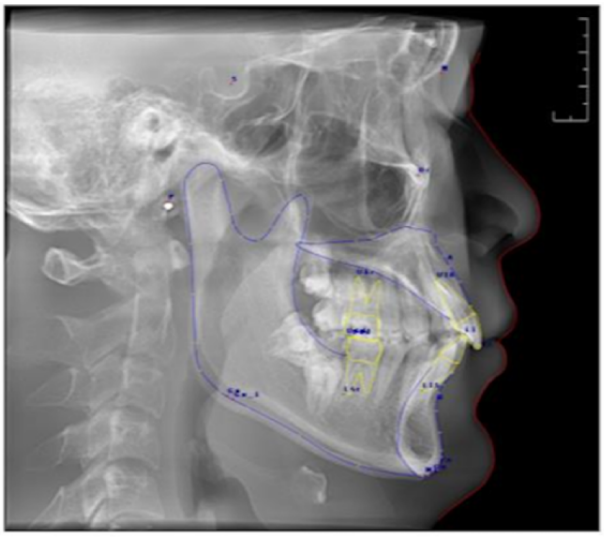
**Figure 3.** Panoramic image (top) showing vertically impacted teeth 18, 28, and 47; a horizontally impacted tooth 48; mesial angulated impaction of tooth 38; and overeruption of teeth 17 and 27 and infraocclusion of tooth 37. Cone beam computed tomography images (bottom) illustrating the spatial relationship of the impacted mandibular second molar (MM2) with respect to the adjacent dentition. (A) Sagittal view (P, posterior; A, anterior) showing an intact mandibular first molar (MM1) distal root surface at the contact point with the MM2 crown; (B) axial view (R, right; lingual, lingual side) showing minimal external resorption of the MM1 distal root at the contact point with the MM2 crown (red arrow); (C) axial view showing no contact and resorption further down the MM1 distal root (red arrow).



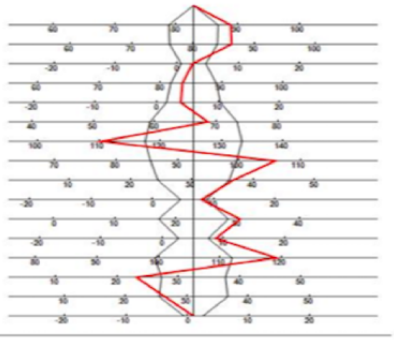



XSL•FO
RenderX

**Figure 4.** Cephalometric analysis according to Chinese norms showing bimaxillary protrusion (increased sella turcica-nasion-point A and sella turcica-nasion-point B), a class I skeletal relationship (point A-nasion-point B), proclined incisors (decreased upper incisor–lower incisor), proclined lower incisors (increased lower incisor–mandibular plane), proclined upper incisors (increased upper incisor-sella turcica-nasion), and a hypodivergent face (decreased mandibular plane–sella turcica-nasion angle).



| 项目名 | 均值 | 标准差 | 测量值 |
|---|---|---|---|
| SNA(dg) | 82.8 | 4 | 88.85 |
| SNB(dg) | 80.1 | 3.9 | 86.31 |
| ANB(dg) | 2.7 | 2 | 2.55 |
| NP-FH(dg) | 85.4 | 3.7 | 83.60 |
| NA-PoA(dg) | 6 | 4.4 | 3.93 |
| Y轴角 | 66.3 | 7.1 | 68.68 |
| U1-L1(dg) | 125.4 | 7.9 | 110.30 |
| L1-MP(dg) | 92.6 | 7 | 105.97 |
| L1-NB(dg) | 30.3 | 5.8 | 36.74 |
| L1-NB(mm) | 6.7 | 2.1 | 8.01 |
| U1-NA(dg) | 22.8 | 5.7 | 30.41 |
| U1-NA(mm) | 5.1 | 2.4 | 8.74 |
| U1-SN(dg) | 105.7 | 6.3 | 119.26 |
| MP-SN(dg) | 32.5 | 5.2 | 23.20 |
| FH-MP(dg) | 31.1 | 5.6 | 26.42 |
| Po-NB(mm) | 1 | 1.5 | 1.04 |



## Treatment Options

Considering the age of the patient, the developmental stage of the impacted MM2 roots and its position relative to the surrounding structures, the type of malocclusion, and the functional importance of the MM2, two following treatment options were suggested: extraction of teeth 38, 48, and the MM2; orthodontic treatment with fixed appliances; and implant replacement of the MM2; or extraction of teeth 38 and 48; attempted extrusion of the MM2 by surgical exposure and orthodontic mechanics, with extraction planned if no movement of the MM2 occurred in 2 to 3 months.

## Treatment Progress

The treatment options were explained to the patient highlighting their pros and cons. For psychosocial and economic reasons, the patient consented to the second treatment option.

Fixed orthodontic appliances were bonded to both jaws to begin tooth alignment and leveling with a 0.014-inch nickel-titanium (NiTi) archwire.

Under local anesthesia, tooth 48 was extracted, and tooth 47 was exposed during the same sitting. A passive lingual button with a traction chain was bonded to its occlusal surface and sutured.

The patient was studying in a distant city, so monthly visits were not feasible for him. Therefore, he was seen again after 5 months. At that visit the archwires were changed in both jaws to 0.016-inch NiTi archwires to finalize the tooth alignment and leveling.

At month 6, the archwires were changed to 0.017×0.025-inch stainless steel archwires on both jaws. A ligature wire in a figure-of-8 configuration was placed in the maxilla. A round loop was then customized using a stainless-steel wire of 0.7 mm and bonded to the occlusal surface of tooth 46 with a distal overhang of about 10 mm. To avoid occlusion imbalance and damage to the temporomandibular joint, a special light-cured resin was bonded to the contralateral maxillary first molar. A vertical force of about 50 to 60 gf was applied by tying an elastic between the round loop and the traction chain bonded to the MM2 occlusal surface. At month 8, the impacted MM2 was visible in the oral cavity (Figure 5), and the same mechanics were maintained until the next follow-up visit.

At month 11, nearly half of the MM2 crown was visible in the mouth. The occlusal traction chain, the loop, and the resin on tooth 26 were removed. A buccal molar tube was bonded to the MM2, and the archwire was changed to a 0.016-inch NiTi on the mandible.

At month 13, the MM2 reached the occlusal level. The archwire was changed to a 0.017×0.025-inch stainless steel on the



XSL•FO
RenderX

mandible for finishing. One month later treatment was completed.

Treatment was successfully completed in 14 months by bringing the tooth into occlusion. Radiographic investigation immediately after treatment revealed sufficient repositioning of the MM2 into the occlusal plane. Axial views showed external root resorption (about 1 mm) on the distolingual surface of the MM1 distal root (Figure 6). Pulp vitality tests of the MM1 and MM2 revealed vital pulp responses and no symptoms were reported by the patient during the course of treatment. The patient was provided with an Essix retainer to be worn for 1.5 years.

**Figure 5.** Photos at month 8. The impacted mandibular second molar was visible in the oral cavity.

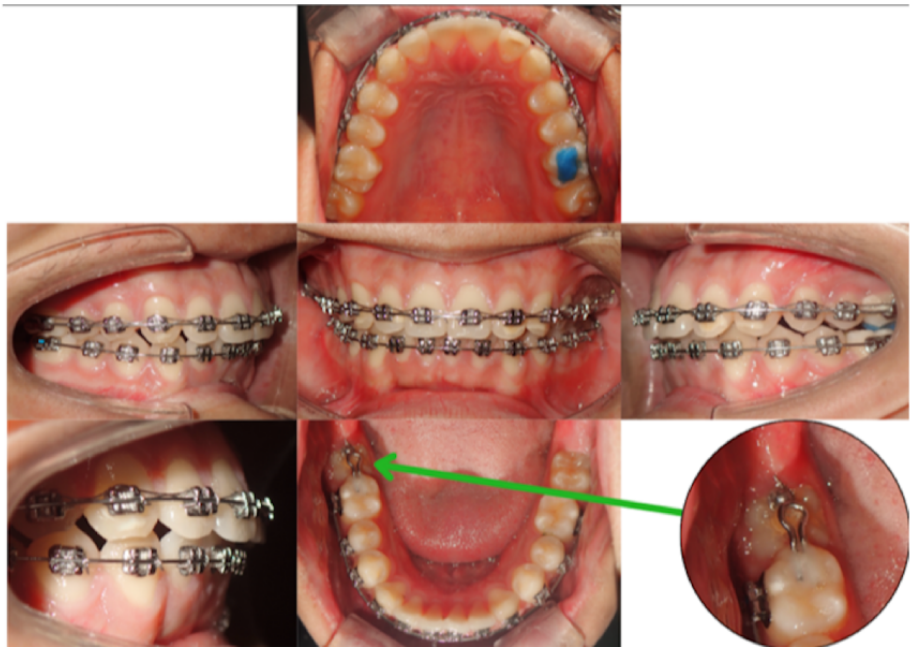
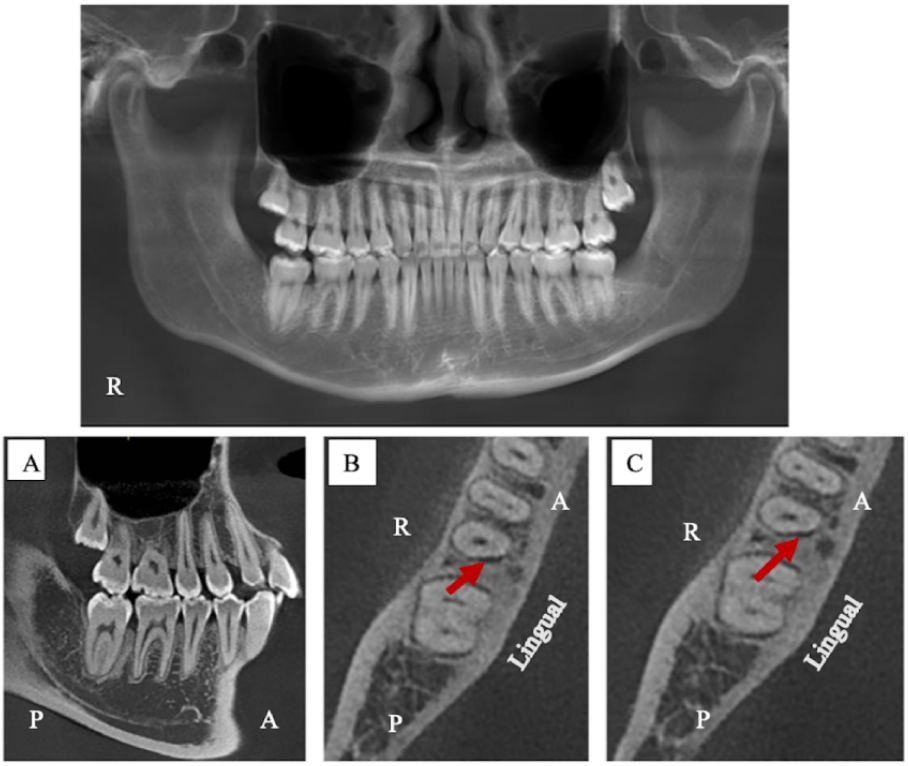


**Figure 6.** Panoramic radiography (top) immediately after treatment showing sufficient repositioning of the mandibular second molar (MM2) into the occlusal plane with intact apical area. Cone beam computed tomography imaging (bottom) immediately after treatment. (A) Sagittal view (P, posterior; A, anterior) showing successful alignment of the MM2 into occlusion; (B and C) axial views (R, right; lingual, lingual side) demonstrating root resorption on the distal root surface of the mandibular firts molar (red arrows).





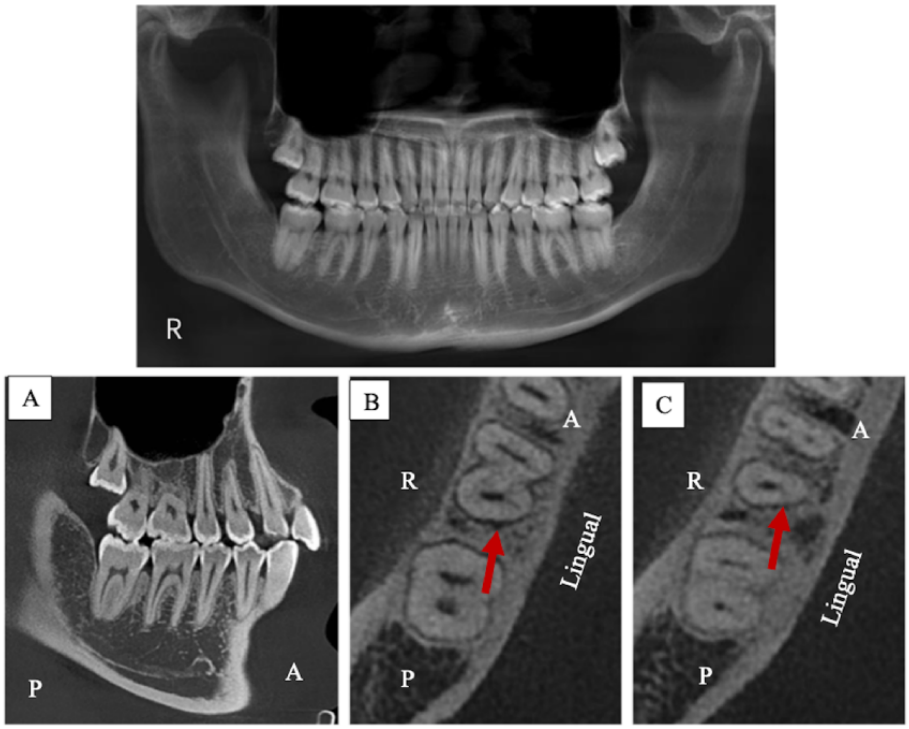

XSL•FO
RenderX

## Follow-Up at 10 Months After Treatment

The patient was recalled for follow-up at 10 months after treatment. No new concerns were reported by the patient. No significant findings were observed upon intraoral examination. Percussion tests in the vertical and horizontal directions were both negative. Pulp vitality tests performed on the MM1 and MM2 indicated vital pulp status. Intraoral photographs (Figure 7) and cone beam computed tomography (CBCT) imaging were carried out. Radiographs showed that the MM1 and MM2 were healthy and free from apical pathologies, and they revealed a reduction in the radiolucent area around the resorption spot to about 0.5 mm (Figure 8).

**Figure 7.** Intraoral photos 10 months after treatment showing good oral hygiene with stable orthodontic treatment results and healthy mandibular first and second molars.

**Figure 8.** Panoramic view (top) at a follow-up visit 10 months after treatment showing healthy mandibular first and second molars (MM1 and MM2) with no signs of periapical pathology. Cone beam computed tomography imaging (bottom) 10 months after treatment. (A) Sagittal view (P, posterior; A, anterior) showing a healthy MM2 free from apical pathology, and an intact distal surface of the MM1 distal root; (B and C) axial views (R, right; lingual, lingual side) demonstrating progressive narrowing of the radiolucent area of root resorption on the distal root surface of the MM1 (red arrows).






XSL•FO
RenderX

## Discussion

This report describes a possible tooth-borne extrusion technique for a vertically impacted MM2 in a 20-year-old adult. Treatment included fixed appliances, extraction of the horizontally impacted tooth 48, and surgical exposure of tooth 47. Tooth 47 became visible at month 8, two months after traction began, reached the occlusal level at month 13, and treatment was completed at month 14. Teeth 46 and 47 retained vital pulp responses, with no reported symptoms or apical pathologies at the 10-month follow-up. The radiolucent area associated with external root resorption of tooth 46 also decreased. These findings demonstrate feasibility in this patient, not comparative effectiveness or wider predictability.

Most published reports concern mesially angulated MM2s; only 8% were vertically oriented in the series by Shpack et al [8]. This explains why existing biomechanical guidance emphasizes uprighting. In our patient, the position of tooth 47 between the distal root of tooth 46 and tooth 48 made local mechanical obstruction a plausible contributor. However, the Angle class II subdivision, 3 mm of anterior crowding, and lingual inclination of teeth 37 and 42 do not prove posterior arch length deficiency or distal drift of tooth 46. Theories linking differential root growth to mesially angulated impaction [9,12] also cannot be directly applied to this vertical case. The etiology therefore remains uncertain rather than attributable to a single demonstrated cause.

CBCT demonstrated the position of tooth 47, its relationship with adjacent structures, and the preexisting resorption on tooth 46. No obvious dilaceration or clear ankylosis-related change was identified, supporting an attempt at orthodontic retrieval. Nevertheless, CBCT cannot definitively exclude ankylosis, and its interpretation is affected by image quality and equipment parameters [1,13]. The planned extraction of tooth 47 if no movement occurred within 2 to 3 months was therefore an important contingency. Subsequent movement and crown emergence argued against ankylosis. Conversely, failure to move would not alone have confirmed it because inadequate force delivery, an unsuitable vector, or appliance problems can also prevent movement [14-16].

Treatment selection considered age, impaction morphology, adjacent root resorption, third molar condition, and patient priorities. Although approximately 14 years has been reported as an optimal treatment age [2,9,17], successful management in adults has been described [18-20]. The alternatives here were extraction of tooth 47 with implant replacement or preservation through exposure and extrusion, with extraction retained as an option if movement failed [21]. Tooth 48 was horizontally impacted and associated with recurrent pericoronitis, whereas tooth 47 was considered functionally worth preserving. The patient's psychosocial and economic preference for avoiding implant treatment informed the shared decision. This justifies the approach in this case but does not establish clinical or economic superiority.

The mechanical requirement was predominantly vertical extrusion rather than correction of severe mesial angulation. Reported superelastic wires, springs, loops, tubes, and miniscrew-supported systems principally address uprighting [6,18-20,22,23]. At month 6, with 0.017×0.025-inch stainless steel archwires in place, a 0.7-mm stainless steel loop with an approximately 10-mm distal extension was bonded to tooth 46. Elastic traction to tooth 47 applied approximately 50 to 60 gf. At month 11, after adequate crown exposure, the traction appliance was replaced by a buccal tube for conventional alignment and finishing. This staged method used routine materials, provided a direct extrusive vector, and avoided a temporary anchorage device (TAD).

The method also had disadvantages. Tooth-borne anchorage created reciprocal forces on tooth 46 and the supporting arch, requiring particular caution because tooth 46 already had root resorption. Although the stainless steel archwires were intended to provide stability, anchorage loss was not measured. Elastic force can decay between visits, and no serial measurements verified the force maintained over time; the reported 50 to 60 gf therefore represents applied rather than proven continuous force. The resin buildup on tooth 26 helped manage occlusion during traction, but its specific temporomandibular joint effect cannot be evaluated from this report. Thus, simplicity and avoidance of a miniscrew must be balanced against anchorage and force-control limitations.

Attendance was another challenge. Because the patient studied in a distant city, a 5-month interval followed appliance placement and exposure, and active traction began at month 6. This limited early review may have extended alignment; it should not be considered a recommended schedule. Moreover, the 14-month total included alignment, extrusion, and finishing, so it cannot support claims that vertical impactions are faster to treat. Immediately after treatment, tooth 46 showed approximately 1 mm of distal root resorption; at 10 months, the corresponding radiolucency was approximately 0.5 mm. Together with vital pulp responses and negative percussion, this suggests a favorable short-term course. However, radiographic narrowing is not histological proof of complete repair, and the respective contributions of impaction, surgery, and traction cannot be separated.

This report is limited by its single-case design, absence of a comparison group, irregular attendance, and short 10-month follow-up. The etiological interpretation is retrospective. Serial force measurements, quantitative anchorage and periodontal assessments, and patient-reported outcomes were not available. CBCT and the response to traction also leave some diagnostic uncertainty. Consequently, long-term stability, complete repair of the adjacent root defect, and durable pulpal and periodontal health remain unconfirmed. Additional reports with standardized measurements and longer follow-up are needed before conclusions about prognosis, treatment duration, or comparative effectiveness can be drawn.

## Clinical Significance

Successful uprighting of a vertically impacted MM2 can be achieved in adults. Vertically impacted MM2s necessitate distinct biomechanical approaches compared with conventional uprighting techniques. Tooth preservation strategies remain viable even when implant alternatives exist.



XSL•FO
RenderX

## Clinical Practice Recommendations

Clinical practice recommendations based on this case are summarized in Table 1.

**Table 1.** Clinical practice recommendations.

| Aspect | Recommendation |
|---|---|
| Diagnosis | 3D imaging should be considered when it is expected to improve assessment of tooth position, root morphology, or adjacent structures. |
| Etiology | A thorough assessment of arch relationships and first molar position, as well as adjacent structures should be carried out in such cases. |
| Ankylosis | Radiological findings should be interpreted together with the clinical response to traction; neither alone resolves every diagnostic uncertainty. |
| Treatment | Orthodontic extrusion may be considered in selected adults; extraction and implant replacement remain alternatives when extrusion is unsuitable or unsuccessful. |
| Mechanics | Vertical extrusion mechanics should be individualized to the tooth position and available anchorage. |
| Follow-up | Posttreatment follow-up should monitor pulp vitality, root resorption, periodontal health, and stability. |

## Conclusions

This case demonstrates that surgical exposure combined with tooth-borne orthodontic extrusion can be a feasible method for preserving a vertically impacted MM2 in a carefully selected adult patient. The main clinical lesson is the importance of matching the mechanics to the impaction pattern. In this patient, treatment required removal of a local obstruction, application of a predominantly vertical extrusive force, and subsequent conventional alignment rather than reliance on uprighting mechanics alone. A customized loop and elastic traction offered a simple option without TADs, but this convenience must be balanced against elastic force decay, reciprocal loading of the anchor teeth, and the need for regular review. Preservation of the natural tooth was consistent with the patient's preferences and avoided implant replacement in this case; it does not establish that this approach is superior, faster, or cost-effective generally. Clinicians considering similar treatment should assess the tooth and adjacent structures carefully, discuss extraction and prosthetic alternatives, retain a contingency for unsuccessful movement, and monitor both the retrieved molar and the adjacent first molar over time. Longer-term reports and comparative case series are required to define prognosis and the circumstances in which this technique is most appropriate.

### Funding

The authors declared that no financial support was received for this work.

### Data Availability

Data used in this report can be obtained from the corresponding author upon reasonable request.

### Generative AI Use

The authors declare that generative AI was used in the preparation of this manuscript. Kimi (Kimi K3.1) was used for proofreading, editing, and language polishing. The authors verified all outputs of the tool and take full responsibility for the final content of the manuscript.

### Authors' Contributions

MHM designed the work and contributed to writing the manuscript. MPP and DDG were responsible for photograph preparation for the manuscript. XG, SLX, and JL were responsible for cone beam computed tomography image acquisition and interpretation. Yue Zhang and Yan Zhang were responsible for study model preparation. CL was responsible for providing guidance, reviewing, and proofreading the manuscript. All authors read and approved the final manuscript.

### Conflicts of Interest

None declared.

### References

1. Barone S, Cevidanes L, Bocchino T, Michelotti A, Borelli M, Giudice A. Mandibular second molar impaction: introducing a novel and validated 3D classification system. BMC Oral Health 2024 Oct 10;24(1):1209. [doi: 10.1186/S12903-024-05006-X]



XSL•FO
RenderX

2. Turley PK. The management of mesially inclined/impacted mandibular permanent second molars. J World Fed Orthod 2020 Oct;9(3S):S45-S53 [FREE Full text] [doi: 10.1016/j.ejwf.2020.09.004] [Medline: 33023732]
3. Fu PS, Wang JC, Wu YM, Huang TK, Chen WC, Tseng YC, et al. Impacted mandibular second molars. Angle Orthod 2012 Jul;82(4):670-675 [FREE Full text] [doi: 10.2319/102111-656.1] [Medline: 22229822]
4. Shapira Y, Finkelstein T, Shpack N, Lai YH, Kuftinec MM, Vardimon A. Mandibular second molar impaction. Part I: genetic traits and characteristics. Am J Orthod Dentofacial Orthop 2011 Jul;140(1):32-37. [doi: 10.1016/j.ajodo.2009.08.034] [Medline: 21724084]
5. Chen HM, Hwang MJ, Chiang CP. Risk factors and treatments for impacted permanent second molars. J Dent Sci 2024 Apr;19(2):1174-1181 [FREE Full text] [doi: 10.1016/j.jds.2023.11.023] [Medline: 38618125]
6. Derton N, Palone M, Siciliani G, Albertini P, Cremonini F, Lombardo L. Resolution of lower second molar impaction through miniscrew-supported biomechanics: a proposal for a simplified classification. Int Orthod 2021 Dec;19(4):697-706. [doi: 10.1016/j.ortho.2021.09.008] [Medline: 34696999]
7. Nemec M, Garzarolli-Thurnlackh G, Lettner S, Nemec-Neuner H, Gahleitner A, Stavropoulos A, et al. Prevalence and characteristics of and risk factors for impacted teeth with ankylosis and replacement resorption – a retrospective, 3D-radiographic assessment. Prog Orthod 2024 Aug 19;25(1):34. [doi: 10.1186/S40510-024-00531-5]
8. Shpack N, Finkelstein T, H Lai Y, M Kuftinec M, Vardimon A, Shapira Y. Mandibular permanent second molar impaction treatment options and outcome. Open J Dent Oral Med 2013;1(1):9-14. [doi: 10.13189/ojdom.2013.010103]
9. Shen HT, Cheng JH, Chen DD. Treatment strategy of mandibular second molar impaction: a literature review. Taiwan J Orthod 2023;35(1). [doi: 10.38209/2708-2636.1329]
10. Ahmed Asif J, Yusuf Noorani T, Khursheed Alam M. Tooth auto-transplantation: an alternative treatment. Bull Tokyo Dent Coll 2017;58(1):41-48 [FREE Full text] [doi: 10.2209/tdcpublication.2016-0011] [Medline: 28381733]
11. Lee TY, Cheng JH, Fang C, Chen DD. Extraction or nonextraction: orthodontic treatment for the complete impaction of second permanent molars. J Dent Sci 2023 Jan;18(1):458-460 [FREE Full text] [doi: 10.1016/j.jds.2022.09.004] [Medline: 36643219]
12. Shapira Y, Lai Y, Schonberger S, Shpack N, Finkelstein T. Mandibular second molar impaction-part II: etiology and role of the third molar. Appl Sci 2022;12(22):11520. [doi: 10.3390/app122211520]
13. Paris M, Trunde F, Bossard D, Farges JC, Coudert JL. L'ankylose dentaire: diagnostic par tomodensitométrie et reconstruction tridimensionnelle [Dental ankylosis diagnosed by CT with tridimensional reconstructions] [Article in French]. J Radiol 2010 Jun;91(6):707-711. [doi: 10.1016/s0221-0363(10)70101-3] [Medline: 20808272]
14. Mubeen S, Seehra J. Failure of eruption of first permanent molar teeth: a diagnostic challenge. J Orthod 2018 Jun;45(2):129-134. [doi: 10.1080/14653125.2018.1462902] [Medline: 29659339]
15. Becker A, Chaushu S. Palatally impacted canines: the case for closed surgical exposure and immediate orthodontic traction. Am J Orthod Dentofacial Orthop 2013 Apr;143(4):451-459. [doi: 10.1016/j.ajodo.2013.02.012] [Medline: 23561404]
16. Becker A, Abramovitz I, Chaushu S. Failure of treatment of impacted canines associated with invasive cervical root resorption. Angle Orthod 2013 Sep;83(5):870-876 [FREE Full text] [doi: 10.2319/090812-716.1] [Medline: 23343004]
17. la Monaca G, Cristalli MP, Pranno N, Galluccio G, Annibali S, Pippi R. First and second permanent molars with failed or delayed eruption: clinical and statistical analyses. Am J Orthod Dentofacial Orthop 2019 Sep;156(3):355-364. [doi: 10.1016/j.ajodo.2018.09.020] [Medline: 31474265]
18. Kang SG, Lim SH, Kim HM, Park SJ, Kim EJ, Park JH. Uprighting a mesially impacted mandibular second molar with scissor bite using simple tubes. J World Fed Orthod 2024 Aug;13(4):189-196. [doi: 10.1016/j.ejwf.2024.02.002] [Medline: 38653656]
19. Bagala F, Maino G, Maino G, Dalessandri D. A superelastic loop for uprighting mesially impacted lower second molars. J Clin Orthod 2019;53(12):726-731. [Medline: 32059224]
20. Chen HH, Chen G, Feng GY, Wang XJ, Xu TM, Su H. Orthodontic treatment of an adult maxillomandibular protrusion case with impacted mandibular second molars using the physiologic anchorage spee-wire system: a case report. Chin J Dent Res 2024 Dec 06;27(4):333-343 [FREE Full text] [doi: 10.3290/j.cjdr.b5860297] [Medline: 39641297]
21. Lee MY, Park JH, Pang KM, Chang NY, Chae JM. Interdisciplinary treatment of a mutilated dentition with multiple missing teeth and horizontal impaction of a mandibular third molar using an orthodontic skeletal anchorage and dental implants-case report. J Esthet Restor Dent 2024 Nov;36(11):1502-1510. [doi: 10.1111/jerd.13281] [Medline: 39012042]
22. Okihara H, Ono T. Correction of bilateral heavily impacted second molar with improved super-elastic nickel-titanium alloy wires. Am J Orthod Dentofacial Orthop 2020 Jul;158(1):114-125 [FREE Full text] [doi: 10.1016/j.ajodo.2019.03.033] [Medline: 32600748]
23. Altieri F, Guarnieri R, Mezio M, Padalino G, Cipollone A, Barbato E, et al. Uprighting impacted mandibular second molar using a skeletal anchorage: a case report. Dent J (Basel) 2020 Nov 18;8(4):129 [FREE Full text] [doi: 10.3390/dj8040129] [Medline: 33218010]


## Abbreviations

**CBCT:** cone beam computed tomography



XSL•FO
RenderX

**MM1:** mandibular first molar
**MM2:** mandibular second molar
**MM3:** mandibular third molar
**NiTi:** nickel-titanium
**TAD:** temporary anchorage device

*Edited by T de Azevedo Cardoso; submitted 20.May.2026; peer-reviewed by G Yordanova, D Rajendraprasad; comments to author 20.Oct.2026; revised version received 05.Nov.2026; accepted 09.Nov.2026; published 26.Nov.2026.*

*Please cite as:*
*Mrisho MH, Lingling C, Pallangyo MP, Li J, Galikano DD, Guo X, Zhang Y, Xiao SL, Zhang Y*
*Orthodontic Management of Vertically Impacted Mandibular Permanent Second Molar in Adults: Case Report*
*Interact J Med Res 2026;15:e101851*
*URL: https://www.i-jmr.org/2026/1/e101851*
*doi:10.2196/101851*
*PMID:*

©Mashaka Hassan Mrisho, Cui Lingling, Mary Paul Pallangyo, Jiaao Li, Deogratius David Galikano, Xinwei Guo, Yue Zhang, Si Long Xiao, Yan Zhang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 26.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# The Shifting Landscape of Natural Health Product Licensing in Canada: The e-Commerce Case of Stem Cell Supplements

Alessandro R Marcon, MA; Michael Jewer, PhD; Timothy Caulfield, LLM

Faculty of Law, Health Law Institute, University of Alberta, Office 470, Edmonton, AB, Canada

**Corresponding Author:**
Alessandro R Marcon, MA
Faculty of Law, Health Law Institute, University of Alberta, Office 470, Edmonton, AB, Canada

## Abstract

Natural health product (NHP) regulations have shifted in Canada, with the aim of reducing "red tape" and enabling market entry. This study assessed how many "stem cell" supplements for sale on Amazon.ca had Health Canada–required natural product number (NPN) licenses compared with vitamin D and "longevity" supplements. It also conducted a marketing overview analysis comparing the Amazon.ca discourse across the 3 product types. Findings show a gap in regulatory licensing compliance between "stem cell" supplements and vitamin D or "longevity" supplements. Marketing discourse also presented potential conflict with regulatory parameters. We discuss the impacts of loosening NHP regulations and the potential for products with weak evidentiary bases to enhance validation through NPN licensing.

***(Interact J Med Res 2026;15:e86747)*** doi:10.2196/86747

**KEYWORDS**

stem cells; regenerative medicine; Amazon; e-commerce; supplements; natural health products; regulation

## Introduction

Scientifically unsupported commercial activity has long accompanied exciting regenerative medicine research [1]. "Stem cell" supplements are now emerging in online marketplaces, including Amazon. In 2024, our research team assessed "stem cell" supplement marketing on Amazon.com, finding inconsistent adherence to Canadian and American regulations [2]. In Canada, supplements that are natural health products (NHPs) require a licensed good manufacturing practices site and product license (natural product number [NPN]) from Health Canada for marketplace entry [3]. Internal analysis on our 2024 study data found that many "stem cell" supplements lacked traceable NPNs in Health Canada's NHP database. In 2025, Health Canada implemented NHP regulatory changes to reduce regulatory "red tape" and facilitate market access [4]. This project's goal was to reassess whether, in July 2026, stem cell supplements on Amazon.ca aligned with NPN license—and broader NHP—regulations.

## Methods

### Overview

In July 2026, we deployed the Apify Amazon Scraper (Apify), a data scraper, to build a dataset of "stem cell supplement" advertisement data for 100 relevant products on Amazon.ca. We also downloaded advertising data from 100 vitamin D and 100 longevity supplements (search: "vitamin D supplement" and "longevity supplement") to act as control and comparison product types. Vitamin D is popular and typically unrelated to stem cells/regenerative medicine, whereas "longevity" supplements may be associated with aging-related health claims similar to stem cell supplement marketing [2]. Downloaded data, held and analyzed on Google Sheets, included over 400 data columns for each product (including labels, product and use descriptions, and details).

Analysis consisted of describing the central advertising themes across the 3 product types (vitamin D, longevity, and stem cell supplements) and determining whether each product had a legitimate NPN (traceable and active) on Health Canada's NHP database [5]. To discern central themes, we deployed a straightforward keyword analysis, modeled on our previous research [2], establishing 6 overarching categories: cell activity, health topic, serious health topic, general effect, general qualifier, and science (Table 1). This analysis provided a general marketing overview and potential distinctions between the 3 product types. To identify NPNs, we searched product advertising data, including images (eg, product labels), and, if found, verified NPN legitimacy. If no product NPN was found in the advertisement data, we searched the NHP database for product or brand name(s). Products marketed for animals or exclusively for skin care were excluded, as were duplicates. The data from products with broken Amazon URLs or products listed as "not currently available" were analyzed but excluded from the results.



XSL•FO
RenderX

**Table .** Health Canada natural product number (NPN) verification and marketing keywords for 3 supplement types on Amazon.ca.

| | Vitamin D (N=78) | Longevity supplements (N=90) | Stem cell supplements (N=80) |
|---|---|---|---|
| Product characteristics | | | |
| NPNs found, n (%) | 78 (100) | 90 (100) | 73 (91.3) |
| Unique NPNs, n (%) | 78 (100) | 86 (95.6) | 70 (87.5) |
| Unique brands, n | 47 | 44 | 56 |
| Marketing keywords by category, n (%) | | | |
| Cell activity | | | |
| Stem cell | 0 (0) | 2 (2.2) | 7 (8.8) |
| Rejuv_ | 0 (0) | 4 (4.4) | 9 (11.3) |
| Regen_ | 1 (1.3) | 3 (3.3) | 6 (7.5) |
| Nicotinamide[a] | 0 (0) | 27 (30) | 27 (33.8) |
| DNA | 2 (2.6) | 8 (8.9) | 7 (8.8) |
| DNA repair | 0 (0) | 6 (6.7) | 5 (6.3) |
| Cell | 10 (12.8) | 49 (54.4) | 58 (72.5) |
| Cellular | 1 (1.3) | 37 (41.1) | 38 (47.5) |
| Health topic | | | |
| Aging | 2 (2.6) | 32 (35.6) | 29 (36.3) |
| Anti-aging | 0 (0) | 14 (15.6) | 17 (21.3) |
| Immun_ | 61 (78.2) | 30 (33.3) | 24 (30) |
| Inflamm_ | 2 (2.6) | 9 (10) | 10 (12.5) |
| Cardiov_ | 11 (14.1) | 11 (12.2) | 10 (12.5) |
| Stress | 6 (7.7) | 24 (26.7) | 20 (25) |
| Oxidative | 0 (0) | 23 (25.6) | 23 (28.8) |
| Energy | 11 (14.1) | 49 (54.4) | 40 (50) |
| Brain related[b] | 7 (9) | 29 (32.2) | 16 (20) |
| Mood | 6 (7.7) | 4 (4.4) | 0 (0) |
| Bone | 76 (97.4) | 6 (6.7) | 6 (7.5) |
| Joint | 7 (9) | 8 (8.9) | 11 (13.8) |
| Heart | 15 (19.2) | 12 (13.3) | 8 (10) |
| Weight loss | 0 (0) | 1 (1.1) | 1 (1.3) |
| Recover | 2 (2.6) | 5 (5.6) | 5 (6.3) |
| Longev_ | 0 (0) | 37 (41.1) | 12 (15) |
| Sleep | 1 (1.3) | 5 (5.6) | 3 (3.8) |
| Gut | 1 (1.3) | 13 (14.4) | 5 (6.3) |
| Toxi_ | 0 (0) | 5 (5.6) | 5 (6.3) |
| Serious health topic | | | |
| Cancer | 1 (1.3) | 0 (0) | 1 (1.3) |
| Asthma | 0 (0) | 0 (0) | 0 (0) |
| Depression | 0 (0) | 0 (0) | 0 (0) |
| Disease | 2 (2.6) | 2 (2.2) | 3 (3.8) |
| Alzh_ | 0 (0) | 0 (0) | 0 (0) |
| Dementia | 0 (0) | 0 (0) | 0 (0) |



XSL•FO
RenderX

| | Vitamin D (N=78) | Longevity supplements (N=90) | Stem cell supplements (N=80) |
|---|---|---|---|
| General effect | | | |
| Heal | 73 (93.6) | 83 (92.2) | 66 (82.5) |
| Repair | 0 (0) | 14 (15.6) | 18 (22.5) |
| Cure | 1 (1.3) | 1 (1.1) | 2 (2.5) |
| Treat | 0 (0) | 1 (1.1) | 3 (3.8) |
| Prevent | 28 (35.9) | 3 (3.3) | 3 (3.8) |
| Diagnose | 0 (0) | 1 (1.1) | 0 (0) |
| Antiox or anti-ox | 6 (7.7) | 45 (50) | 43 (53.8) |
| Detox | 0 (0) | 6 (6.7) | 6 (7.5) |
| Cleanse | 0 (0) | 3 (3.3) | 4 (5) |
| Support | 72 (92.3) | 84 (93.3) | 71 (88.8) |
| Boost | 17 (21.8) | 34 (37.8) | 33 (41.3) |
| Strengthen | 8 (10.3) | 1 (1.1) | 3 (3.8) |
| General qualifier | | | |
| Trust | 15 (19.2) | 16 (17.8) | 15 (18.8) |
| Super | 7 (9) | 14 (15.6) | 13 (16.3) |
| Good | 18 (23.1) | 25 (27.8) | 14 (17.5) |
| Pure | 28 (35.9) | 44 (48.9) | 27 (33.8) |
| Wellness | 17 (21.8) | 49 (54.4) | 25 (31.3) |
| Natural | 45 (57.7) | 66 (73.3) | 51 (63.8) |
| Science related | | | |
| Science | 3 (3.8) | 2 (2.2) | 11 (13.8) |
| Scientific | 0 (0) | 5 (5.6) | 8 (10) |
| Scientist | 0 (0) | 0 (0) | 1 (1.3) |
| Research | 5 (6.4) | 15 (16.7) | 13 (16.3) |
| Proven | 1 (1.3) | 6 (6.7) | 6 (7.5) |
| Evidence | 2 (2.6) | 1 (1.1) | 2 (2.5) |
| Any science-related[c] | 8 (10.3) | 20 (22.2) | 28 (35) |
| Tested | 21 (26.9) | 38 (42.2) | 31 (38.8) |

[a]Includes “NAD+,” “NAD,” “NMN,” and “nicotinamide.”

[b]Includes “brain,” “mental,” “cognit_,” and “memory.”

[c]Refers to all listed terms excluding “tested.”

### Ethical Considerations

Ethics approval and informed consent were not required as all Amazon ads are publicly visible. No product, brand, or company names appear in this manuscript.

## Results

Nearly all supplements had legitimate NPNs. All 78 vitamin D products and all 90 longevity products had legitimate NPNs. Of the 80 stem cell supplements (56 unique brands), 73 (91%) had NPNs. The 7 (9%) products lacking NPNs came from 3 unique brands and explicitly mentioned “stem cell” in their Amazon.ca marketing. Of these 7 products, 2 were available for purchase but showed a discontinued NPN in Health Canada’s database. An additional 2 stem cell products and 1 longevity product did not produce a legitimate NPN but were excluded for having broken URLs or being unavailable for purchase.

NHP companies selling vitamin D, longevity, and stem cell supplements presented different themes in the advertising discourse (Table 1). Compared to vitamin D products, stem cell and longevity supplements presented greater focus on cell activity and covered a much greater range of health topics (eg, energy, aging, immunity, stress, oxidative, and brain related). Vitamin D products focused almost exclusively on immunity and bone health. Stem cell products presented cell activity with greater frequency than longevity supplements. None of the



XSL•FO
RenderX

product types mentioned serious health topics with frequency. The effect descriptors “heal,” “support,” and “boost” were common across all product types, but “cure” and “treat” were scant. The general qualifiers “pure,” “wellness,” and “natural” had a similarly prominent presence across the 3 product types. Stem cell products had the highest presence of science-related discourse (n=28, 35%) compared to longevity (n=20, 22%) and vitamin D (n=8, 10%) (Table 1).

## Discussion

The overwhelming majority of vitamin D, longevity, and stem cell supplements on Amazon.ca in this cross-sectional sample analysis had traceable NPNs in Health Canada’s NHP database. Only stem cell products presented cases of regulatory NPN noncompliance (7/80, 9%). Demonstrating overall regulatory compliance, the marketing for these products scarcely mentioned serious health topics (eg, cancer, dementia) and causal modes of action (eg, cure, treat), which could constitute NHP regulatory breaches [6]. Issues persist, however, about the validation of questionable NHPs by Health Canada’s current regulations and whether obtaining a Health Canada–certified NPN may bolster product credibility.

NHPs certified by Health Canada should not have free rein for deceptive marketing. Both the stem cell and longevity supplements’ marketing in this study frequently mentioned cell activity and a range of health topics. For example, nicotinamide, which is mentioned in approximately 30% of stem cell supplement marketing and 30% of longevity supplement marketing, remains therapeutically speculative despite preclinical development [7]. Vitamin D supplementation also has inconclusive health benefits for nondeficient patients [8]. Canadian NHP regulations restrict marketing that distorts the evidentiary base related to health claims [2]. This project’s broad marketing analysis found science-related discourse in 35% of the stem cell supplement ads and 22% of the longevity supplement ads. Scientific discourse typically works toward promoting evidence around product reliability or efficacy [2]. Notably, stem cell and longevity supplements have a more speculative evidentiary base but more frequently presented appeals to science.

Health Canada’s recent loosening of NHP regulations may be increasing the e-commerce market availability of NHPs, including on Amazon.ca. Correspondingly, the nature of effective surveillance (eg, automated, licensing, risk-based targeting) and enforcement may be fluctuating significantly. The relatively high levels of regulatory compliance shown in this research suggest an opportunity for increasing regulatory pressure around deceptive or overexaggerated product promotion, especially given recent industry-favorable measures to facilitate market access. Ongoing and diligent NHP marketplace surveillance is needed to track the emergence of new products (where compliance may be low) and the corresponding marketing used to promote products with limited, circumstantial, or nonexistent health value for Canadian consumers.

### Acknowledgments

The authors thank Tuvini de Silva for her coding contributions and Robyn Hyde-Law for project management and administrative assistance.

### Funding

This research was funded by the Stem Cell Network under the projects Law, Public Policy and Social License for Next-Generation Regenerative Medicine (grant number 917390), and Regenerative Medicine, Wellness Influencers, and Scienceploitation (grant number IE-C5R1-1). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

### Data Availability

Data have been uploaded to a publicly accessible OSF repository [9].

### Authors' Contributions

ARM and TC conceived the study. MJ collected the data. ARM and MJ performed the analysis. ARM and MJ wrote the manuscript. ARM, MJ, and TC approved the final version.

### Conflicts of Interest

None declared.

### References

1. Turner L, Wang JC, Martinez JR Jr, Najjar S, Rajapaksha Arachchilage T, Sahrai V. US businesses engaged in direct-to-consumer marketing of perinatal stem cell interventions following the Food and Drug Administration’s enforcement discretion era. Cytotherapy 2024 Apr;26(4):393-403. [doi: 10.1016/j.jcyt.2024.01.001] [Medline: 38340106]



XSL•FO
RenderX

2. Marcon AR, Zenone M, Boniface V, Sigfstead S, Murdoch B, Caulfield T. The marketing of “stem cell” supplements on Amazon.com: Assessing alignment with regulatory frameworks in the United States and Canada. Stem Cell Reports 2025 Nov 11;20(11):102675. [doi: 10.1016/j.stemcr.2025.102675] [Medline: 41072411]
3. Natural health product regulation in Canada: overview. Government of Canada. 2026. URL: https://www.canada.ca/en/health-canada/services/drugs-health-products/natural-non-prescription/regulation.html [accessed 2026-09-14]
4. Public Health Agency of Canada, Health Canada. Health Canada and the Public Health Agency of Canada’s report on red tape reduction.: Government of Canada; 2025. URL: https://www.canada.ca/en/services/health/publications/health-system-services/report-red-tape-reduction.html#a6 [accessed 2026-09-18]
5. Health Canada. Search licensed natural health products. Government of Canada. 2024. URL: https://health-products.canada.ca/lnhpd-bdpsnh/?lang=eng [accessed 2026-09-18]
6. Health Canada. Guidance document: Schedule A and Section 3 of the Food and Drugs Act. Government of Canada. 2003. URL: https://www.canada.ca/en/health-canada/services/drugs-health-products/drug-products/applications-submissions/guidance-documents/guidance-document-schedule-section-3-food-drugs-act.html [accessed 2026-09-18]
7. Freeberg KA, Udovich CC, Martens CR, Seals DR, Craighead DH. Dietary supplementation with NAD+-boosting compounds in humans: current knowledge and future directions. J Gerontol A Biol Sci Med Sci 2023 Dec 1;78(12):2435-2448. [doi: 10.1093/gerona/glad106] [Medline: 37068054]
8. Bouillon R, Manousaki D, Rosen C, Trajanoska K, Rivadeneira F, Richards JB. The health effects of vitamin D supplementation: evidence from human studies. Nat Rev Endocrinol 2022 Feb;18(2):96-110. [doi: 10.1038/s41574-021-00593-z] [Medline: 34815552]
9. Marcon AR. Supplements for sale on Amazon.ca (June 27, 2026). OSF. URL: https://osf.io/ka6z8/overview [accessed 2026-09-18]

## Abbreviations

**NHP:** natural health product
**NPN:** natural product number

*Edited by M Balcarras; submitted 29.Oct.2025; peer-reviewed by L Ikonomou, Y Chu; revised version received 16.Aug.2026; accepted 03.Sep.2026; published 28.Sep.2026.*

*Please cite as:*
*Marcon AR, Jewer M, Caulfield T*
*The Shifting Landscape of Natural Health Product Licensing in Canada: The e-Commerce Case of Stem Cell Supplements*
*Interact J Med Res 2026;15:e86747*
*URL: https://www.i-jmr.org/2026/1/e86747*
*doi:10.2196/86747*

© Alessandro R Marcon, Michael Jewer, Timothy Caulfield. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 28.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Sustained Health Care–Associated Infection Reduction Among Electronic Hand Hygiene Monitoring System Clients

Gabrielle Genovesi, MPH

BioVigil Technologies, LLC., 924 N. Main Street, Suite 2, Ann Arbor, MI, United States

**Corresponding Author:**
Gabrielle Genovesi, MPH
BioVigil Technologies, LLC., 924 N. Main Street, Suite 2, Ann Arbor, MI, United States

## Abstract

This multifacility retrospective analysis of 25 hospitals using BioVigil Technologies' electronic hand hygiene monitoring system demonstrated a mean of 46.95% (95% CI 35.5% - 58.4%) sustained reductions in health care–associated infections in 2024 compared to pre-implementation baselines, highlighting the clinical and operational value of automated hand hygiene monitoring at scale.

***(Interact J Med Res 2026;15:e101905)*** doi:10.2196/101905

**KEYWORDS**

electronic hand hygiene monitoring; infection reduction; clinical outcomes; patient safety; hand hygiene compliance; CMS quality reporting; health care–associated infections; Centers for Medicare & Medicaid Services

## Introduction

Health care–associated infections (HAIs) remain a challenge for health care facilities worldwide, contributing to patient morbidity, mortality, and substantial economic burden. In the United States, approximately 1 in 31 hospitalized patients acquires at least 1 HAI, at a cost of billions annually [1]. Despite longstanding evidence-based guidance, sustained improvements in infection rates have proven difficult across diverse health care settings.

Electronic hand hygiene monitoring (EHHM) systems offer a promising approach to enhancing adherence to infection prevention protocols while providing robust data for operational decision-making. By delivering real-time feedback to health care workers and tracking performance at the unit and facility level, these systems can identify and correct lapses in hand hygiene behavior, a primary contributor to HAIs [2]. Prior studies demonstrated that automated monitoring improves compliance and reduces infection rates, though reported effects vary by system design. A recent meta-analysis found EHHM increased compliance approximately 1.6-fold and was associated with markedly lower HAI rates, while trials of group-level dispenser-counting and proximity-based badge systems have shown more modest or inconsistent gains; long-term, multifacility analyses remain limited [3].

This analysis evaluates HAI performance among 25 BioVigil client facilities using publicly reported Centers for Medicare & Medicaid Services (CMS) data. Differing from group-level dispenser counters or proximity-based badges, BioVigil's EHHM technology chemically verifies hand sanitization; confirms soap-and-water wash duration; and its color-coded badge provides real-time feedback through visual, vibrational, and audible cues, supporting cross-contamination prevention and individual-level accountability. By comparing pre- and post-implementation periods, we aim to quantify the clinical and economic impact of EHHM across diverse hospital types and sizes, providing insight into strategies for infection prevention improvement.

## Methods

### Study Design

Facilities were included if BioVigil's EHHM was implemented in 2023 or earlier, CMS HAI data were complete for 2024 and the year preceding implementation, and CMS reporting was available at the individual facility level. These criteria yielded 25 facilities.

The facilities included 16 short-term acute care hospitals, 1 long-term acute care hospital, seven critical access hospitals, and one Veterans Affairs hospital, collectively comprising 4198 beds and 284 clinical units. CMS-reported HAIs for short-term acute care hospitals included catheter-associated urinary tract infections, central line-associated bloodstream infections, methicillin-resistant *Staphylococcus aureus*, surgical site infections, and *Clostridioides difficile* infections; long-term acute care, critical access, and Veterans Affairs hospitals reported catheter-associated urinary tract infections, central line-associated bloodstream infections, and *Clostridioides difficile* infections. Earliest full EHHM implementation occurred in 2016 (2 facilities) and most recently in 2023 (3 facilities).

This retrospective, observational analysis evaluated HAI outcomes aggregated at the hospital level. Percent change from



XSL•FO
RenderX

baseline to 2024 was calculated for each facility and across the study population. Descriptive statistics, including mean, median, and 95% CIs, characterized sustained changes in infection counts. EHHM operational data from 2024 were also analyzed to quantify hand hygiene opportunities, real-time corrective reminders, and overall compliance rates. All data were publicly available CMS reports or deidentified operational records; no patient-level data were used.

### Ethical Considerations

This study utilized only deidentified, publicly available datasets from the CMS and summarized medical device usage counts. Because no individual-level or identifiable patient information was accessed or analyzed, this research did not constitute human subjects research, and institutional review board review was not required.

## Results

In 2024, BioVigil's EHHM system captured 67,663,570 hand hygiene opportunities across all 25 facilities. During this period, 1,112,148 potential cross-contamination events were corrected in real time through badge-based reminders. This contributed to an average hand hygiene compliance rate of 91.25% across the study population.

Of the 25 facilities, 22 demonstrated reductions in total HAI counts compared to their pre-implementation baseline. Four facilities achieved net-zero HAIs in 2024, representing a 100% reduction. One facility exhibited no change, and 2 experienced minimal increases of 2 and 4 cases, respectively—both from unusually low baseline years that likely underrepresent each facility's true pre-EHHM infection burden (Table 1).

**Table .** Facility characteristics and health care–associated infection (HAI) reductions in 2024.

| Faculty name | Facility type | Approximate bed count (rounded to nearest 25) | BioVigil contract start year | HAI percent decrease in 2024 |
|---|---|---|---|---|
| Hospital 1 | Critical access | 25 | 2023 | 100 |
| Hospital 3 | Critical access | 25 | 2018 | 100 |
| Hospital 10 | Short-term acute care | 50 | 2021 | 100 |
| Hospital 15 | Critical access | 25 | 2021 | 100 |
| Hospital 19 | Short-term acute care | 150 | 2023 | 82.35 |
| Hospital 16 | Short-term acute care | 50 | 2021 | 75 |
| Hospital 21 | Short-term acute care | 375 | 2021 | 71.43 |
| Hospital 24 | Veteran's Affairs | 350 | 2023 | 66.67 |
| Hospital 22 | Short-term acute care | 125 | 2022 | 64.71 |
| Hospital 25 | Short-term acute care | 300 | 2014 | 64.06 |
| Hospital 23 | Short-term acute care | 300 | 2018 | 55.71 |
| Hospital 18 | Long-term acute care | 275 | 2020 | 54.55 |
| Hospital 7 | Critical access | 25 | 2021 | 50 |
| Hospital 20 | Short-term acute care | 75 | 2016 | 46.15 |
| Hospital 8 | Critical access | 25 | 2021 | 44.44 |
| Hospital 2 | Short-term acute care | 150 | 2021 | 39.13 |
| Hospital 4 | Short-term acute care | 750 | 2020 | 37.33 |
| Hospital 5 | Short-term acute care | 175 | 2017 | 24.32 |
| Hospital 11 | Short-term acute care | 100 | 2021 | 20 |
| Hospital 6 | Short-term acute care | 300 | 2020 | 17.46 |
| Hospital 9 | Short-term acute care | 200 | 2021 | 13.04 |
| Hospital 13 | Short-term acute care | 150 | 2021 | 11.11 |
| Hospital 12 | Critical access | 25 | 2021 | 0 |
| Hospital 17 | Short-term acute care | 460 | 2019 | –13.79 |
| Hospital 14 | Critical access | 25 | 2021 | –50 |

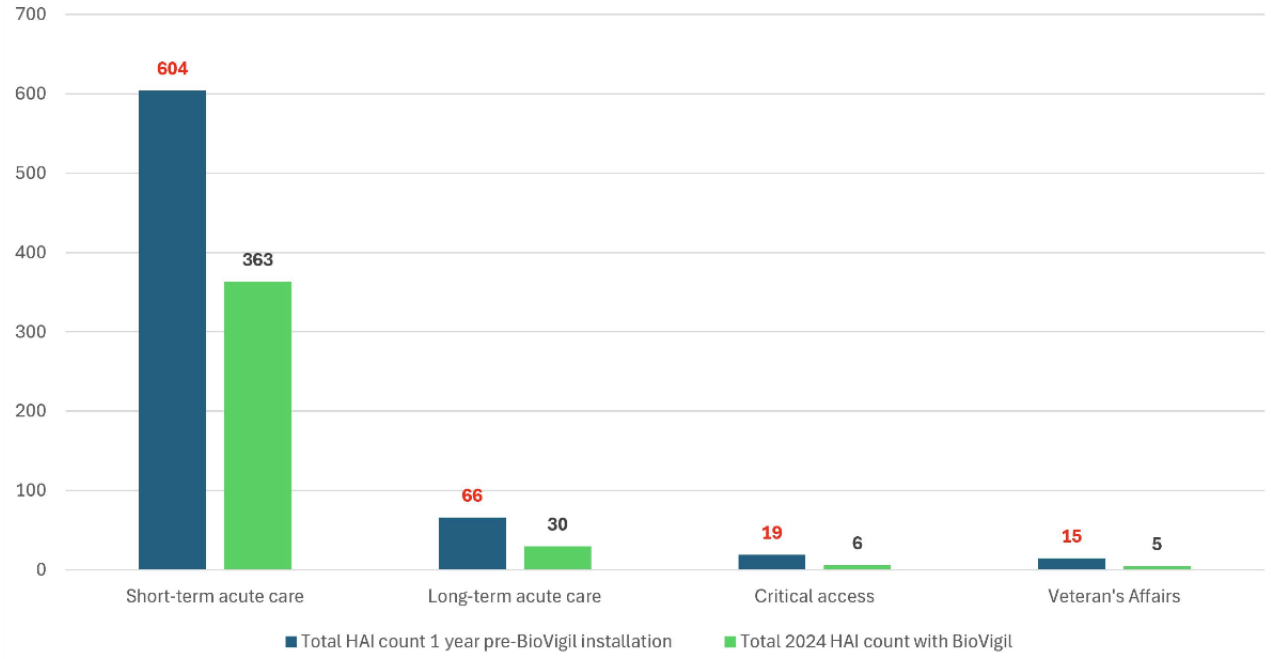
Across all 25 facilities, the mean sustained HAI reduction was 46.95% (95% CI 35.5% - 58.4%), closely aligning with the median of 50% (IQR 30.83%-73.22%) (Figure 1). Among the 22 improving facilities, the mean reduction was 56.25% (95% CI 43.2% - 69.3%) with a median of 55.13% (IQR 37.78%-74.11%), collectively representing 300 fewer CMS-reported HAI cases in 2024. For context, national National Healthcare Safety Network data from over 38,000 US health



XSL•FO
RenderX

care facilities showed year-over-year declines of only 2% - 11% across the 6 CMS-reportable HAI categories during the same period [4].

**Figure 1.** Health care–associated infection (HAI) annual counts pre–BioVigil electronic hand hygiene monitoring (EHHMS) installation versus the 2024 performance with EHHMS (by hospital type; 22 facilities).




Financial and operational efficiencies were also observed, estimated using external benchmark averages. At a mean cost of $29,412 per HAI case [5], the observed reduction corresponds to approximately $8.82 million in avoided costs, excluding extended length of stay, readmissions, and CMS value-based purchasing penalties. Additionally, 16 of 25 facilities reported hand hygiene data to the Leapfrog Group; benchmark labor estimates suggest exclusive reliance on manual observation would have required approximately 136,620 staff hours and $5.74 million in labor costs in 2024 [6].

## Discussion

This analysis demonstrates that the large majority of BioVigil Technologies' client facilities achieved meaningful, sustained reductions in HAIs in 2024. These findings extend prior evidence of EHHM's effectiveness by documenting outcomes across a geographically and structurally diverse portfolio of 25 facilities over extended post-implementation periods. This is an analytical scope not well represented in existing literature, which has largely focused on single-site studies or short-term compliance metrics [3].

While prior work has established that automated monitoring improves adherence, this analysis suggests that sustained, high-level compliance translates into durable reductions in reported infection counts. This suggests EHHM may promote lasting behavioral change beyond attributable observation effects. From clinical and policy perspectives, these results reinforce the case for broader institutional investment in scalable, automated monitoring as a complement to existing infection prevention infrastructure.

This analysis has a few limitations to consider. First, as a retrospective, observational study without a randomized control group, causal attribution between EHHM implementation and HAI reductions cannot be definitively established, as additional infection prevention initiatives may have contributed. Second, baseline comparisons relied on a single pre-implementation year, which may not fully capture natural year-to-year variability, particularly in smaller facilities with low baseline counts. Third, the study population was limited to BioVigil clients that met the inclusion criteria, which may introduce selection bias and limit generalizability.

In conclusion, the majority of BioVigil client facilities experienced substantial HAI reductions in 2024 relative to pre-implementation baselines. This is regardless of system use duration. These outcomes suggest EHHM is an effective, sustainable, data-driven strategy to strengthen infection prevention performance, improve patient safety, and enhance health care efficiency.

### Acknowledgments

The author wishes to acknowledge the Centers for Medicare & Medicaid Services (CMS) for making hospital-acquired infections data publicly available, which served as the foundation for this work. The author also states that generative AI (Claude) was used solely to review the structure of this Research Letter for conformity with journal requirements.



XSL•FO
RenderX

## Funding

The author declared no financial support was received for this work.

## Data Availability

The data that support the findings of this study are available from the author upon reasonable request.

## Authors' Contributions

GG confirms being the sole contributor of this work and was responsible for the conceptualization, methodology, formal analysis, writing (original draft), and writing (review and editing).

## Conflicts of Interest

The author is an employee of BioVigil Technologies, the developer and vendor of the electronic hand hygiene monitoring system evaluated in this analysis.

## References

1. Healthcare-associated infections (HAIs). Centers for Disease Control and Prevention. URL: https://www.cdc.gov/healthcare-associated-infections/index.html [accessed 2026-02-24]
2. Marra AR, Edmond MB. New technologies to monitor healthcare worker hand hygiene. Clin Microbiol Infect 2014 Jan;20(1):29-33. [doi: 10.1111/1469-0691.12458] [Medline: 24245809]
3. Boyce JM, Laughman JA, Ader MH, Wagner PT, Parker AE, Arbogast JW. Impact of an automated hand hygiene monitoring system and additional promotional activities on hand hygiene performance rates and healthcare-associated infections. Infect Control Hosp Epidemiol 2019 Jul;40(7):741-747. [doi: 10.1017/ice.2019.77] [Medline: 31106714]
4. Current HAI progress report: 2024 National and State Healthcare-Associated Infections (HAI) Progress Report. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/healthcare-associated-infections/php/data/progress-report.html [accessed 2026-02-24]
5. Nelson R, Ashby L, Bennett M, et al. 2023. Excess cost of healthcare-associated infections in the Central Texas Veterans Health Care System. Open Forum Infect Dis 2022 Dec 15;9(Supplement_2). [doi: 10.1093/ofid/ofac492.1647]
6. Alper P. Hand hygiene monitoring: budget for 200 direct observations. Medline. 2022 Apr. URL: https://www.medline.com/strategies/infection-prevention/hand-hygiene-observation-whats-the-budget-to-meet-the-leapfrog-group-standard/ [accessed 2026-02-24]

## Abbreviations

**CMS:** Centers for Medicare & Medicaid Services
**EHHM:** electronic hand hygiene monitoring
**HAI:** health care–associated infection

*Edited by M Balcarras; submitted 20.May.2026; peer-reviewed by G Ghssein, NA Amoh-Asante, O Ayilara; revised version received 12.Aug.2026; accepted 13.Aug.2026; published 24.Aug.2026.*

*Please cite as:*
*Genovesi G*
*Sustained Health Care–Associated Infection Reduction Among Electronic Hand Hygiene Monitoring System Clients*
*Interact J Med Res 2026;15:e101905*
*URL: https://www.i-jmr.org/2026/1/e101905*
*doi:10.2196/101905*

© Gabrielle Genovesi. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 24.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Within-Provider Variation in Prices for Commonly Utilized Services in Office Setting: Cross-Sectional Study

Yuvraj Pathak*, PhD; David Muhlestein*, PhD, JD

Simple Healthcare, 465 N Carolina Run, Sanford, FL, United States

*all authors contributed equally

**Corresponding Author:**
Yuvraj Pathak, PhD
Simple Healthcare, 465 N Carolina Run, Sanford, FL, United States

## Abstract

Our study uses the latest Transparency in Coverage data to show that there is wide variation in health care prices for some of the most commonly utilized services in the United States. The findings demonstrate that there is an opportunity for employers, policymakers, and other stakeholders to curb health care spending by choosing cost-efficient health care networks and providers.

***(Interact J Med Res 2026;15:e96883)*** doi:10.2196/96883

**KEYWORDS**

healthcare prices; price transparency; cost of care; transparency in coverage; healthcare price variation

## Introduction

Health care prices in the United States have historically been hidden from the public hindering meaningful analysis of health care markets and prices. This matters because prices impact access to care and are a major driver of high health care spending [1,2].

The information landscape began to change with first price transparency rule, Hospital Price Transparency (HPT) in 2021 which required hospitals to disclose prices of a select set of services. Then, in 2022, the Transparency in Coverage (TiC) rule began to require insurers to release negotiated rates for all covered services.

These datasets have proven valuable in shedding light on variation in health care prices. HPT data has been used to demonstrate large variation in prices for urological procedures [3], and radiology services [4], among others. Recent studies have begun to exploit more comprehensive TiC data and show wide variation in prices for inpatient and outpatient services [5], within hospitals for commonly provided services [6], among others.

We assess the within-provider variation in prices for the top 20 most commonly services in an office setting and find that the average difference between the minimum and maximum rates is 40% of the national Medicare rate. Our results complement Wang et al (2024) study which examines within-hospital price variation for inpatient and outpatient services and finds that difference between minimum and maximum prices negotiated by hospitals to be 86% and 222% of Medicare rates, respectively [6].

## Methods

We downloaded the machine-readable files for the largest national payers (Cigna, Aetna, United Healthcare, various Blue Cross and Blue Shield plans) for the third quarter of 2025.

These files contain information on negotiated rates for all covered services along with identifying information on provider tax identification number (TIN), payer name and plan type.

We removed duplicates and ‘ghost rates’—negotiated rates corresponding to a provider for a service they are unlikely to ever bill for,—which are common in TiC data following the established methodology from published literature [7].

This analysis focuses on the top twenty most commonly utilized HCPCS ((Healthcare Common Procedure Coding System) codes in an office setting (determined based on national Medicare data and Colorado and New Hampshire’s All-Payer Claims databases). To limit the influence of outliers, we exclude rates below the 1st percentile and above the 99th percentile for each billing code and restrict the sample to billing code–TIN combinations with at least two negotiated rates.

The analysis data consists of approximately 5.6 million negotiated rates for 20 HCPCS codes in an office place of service. Table 1 shows the distribution of negotiated rates by billing codes.



XSL•FO
RenderX

**Table .** Percentiles and counts of negotiated rates by billing codes.

| Billing code | Billing code description | Percentiles ($) | | | | | Standard deviation | Skewness | No. of negotiated rates |
|---|---|---|---|---|---|---|---|---|---|
| | | 10th | 25th | 50th | 75th | 90th | | | |
| 36415 | Collection of venous blood by venipuncture | $1.80 | $2.91 | $3.00 | $5.50 | $8.21 | 2.52 | 1.16 | 252,841 |
| 80048 | Basic metabolic panel (BMP) | $4.23 | $5.19 | $6.77 | $9.31 | $12.82 | 3.56 | 1.44 | 129,614 |
| 80053 | Comprehensive metabolic panel (CMP) | $5.08 | $6.04 | $8.12 | $11.49 | $15.65 | 4.46 | 1.42 | 136,811 |
| 80061 | Lipid panel | $5.62 | $7.74 | $10.71 | $15.00 | $20.95 | 6.34 | 1.43 | 175,161 |
| 83036 | Hemoglobin $A1_c$ test | $4.67 | $5.70 | $7.77 | $10.85 | $14.86 | 4.24 | 1.42 | 183,042 |
| 84443 | Thyroid stimulating hormone (TSH) test | $7.06 | $9.65 | $13.44 | $18.28 | $25.36 | 7.49 | 1.34 | 148,063 |
| 85025 | Complete blood count (CBC) with differential | $4.63 | $5.44 | $6.29 | $8.70 | $11.96 | 3.15 | 1.47 | 187,763 |
| 85610 | Prothrombin time (PT/INR) | $2.61 | $3.00 | $4.29 | $5.50 | $6.93 | 1.72 | 1.22 | 145,974 |
| 88305 | Surgical pathology examination, moderate complexity (eg, biopsy tissue exam) | $40.94 | $53.50 | $71.84 | $102.59 | $131.49 | 41.50 | 1.43 | 42,117 |
| 90471 | Immunization administration (1 vaccine) | $12.00 | $14.72 | $19.63 | $26.80 | $33.82 | 9.01 | 1.02 | 211,824 |
| 93010 | Electrocardiogram (ECG/EKG) interpretation and report | $7.79 | $9.36 | $12.16 | $14.94 | $17.56 | 4.19 | 1.09 | 88,739 |
| 97110 | Therapeutic exercises (physical therapy), 15 minutes | $20.80 | $24.64 | $29.66 | $34.99 | $45.91 | 10.72 | 1.61 | 215,203 |
| 97112 | Neuromuscular reeducation, 15 minutes | $23.10 | $27.31 | $32.49 | $38.96 | $50.77 | 12.04 | 1.58 | 210,790 |
| 97140 | Manual therapy techniques (eg, mobilization), 15 minutes | $19.14 | $22.55 | $27.24 | $32.48 | $42.59 | 10.20 | 1.62 | 218,521 |



XSL•FO
RenderX

| Billing code | Billing code description | Percentiles ($) | | | | | Standard deviation | Skewness | No. of negotiated rates |
|---|---|---|---|---|---|---|---|---|---|
| | | 10th | 25th | 50th | 75th | 90th | | | |
| 97530 | Therapeutic activities, 15 minutes | $24.93 | $29.71 | $36.50 | $43.32 | $56.53 | 13.60 | 1.46 | 145,810 |
| 98941 | Chiropractic manipulation, 3 - 4 spinal regions | $24.43 | $30.65 | $40.53 | $54.19 | $66.18 | 16.63 | 0.77 | 55,689 |
| 99203 | New patient office visit, moderate complexity | $73.50 | $86.70 | $107.86 | $130.09 | $163.66 | 36.60 | 1.08 | 717,830 |
| 99212 | Established patient office visit, low complexity | $28.66 | $36.22 | $46.53 | $58.03 | $72.91 | 17.42 | 0.89 | 708,342 |
| 99213 | Established patient office visit, low-moderate complexity | $52.06 | $61.28 | $76.02 | $94.90 | $118.35 | 26.58 | 1.01 | 808,269 |
| 99214 | Established patient office visit, moderate complexity | $75.95 | $90.31 | $110.02 | $138.66 | $172.76 | 38.73 | 1.03 | 805,806 |

We first calculate the mean, median, minimum, and maximum rates at the provider TINs and billing code level. These represent the calculated measures of prices a provider is willing to accept across insurers for a billing code.

We then take the average of these estimates across provider TINs at the billing code level to calculate the mean, median, minimum, and maximum rate a provider is willing to accept for a billing code on average.

We define the range of rates as the difference between the average minimum and average maximum rates divided by the national Medicare rate for each billing code.

### Ethical Considerations

This study is exempt from ethics review, as all data used is publicly available. The study uses data on health care prices released by insurers and does not involve any human subjects. It is thus exempt from institutional review board.

## Results

Table 2 shows the average mean, median, minimum, and maximum –negotiated rates for the top 20 billing codes. All four summary statistics are unweighted averages and represent the average of the respective summary statistics across all provider TINs at the billing code level. Out of the 20 billing codes, difference between the maximum and minimum rate is greater than or equal to 50% of the national Medicare rate for four billing codes - 80048 (Basic Metabolic Panel), 80061 (Lipid Panel), 85610 (Prothrombin Time), and 88305 (Surgical Pathology).



XSL•FO
RenderX

**Table .** Average of the mean, median, minimum, maximum values, and estimated range of negotiated rates, by billing code.

| Billing code | Billing code description | Mean (averaged over provider TINs)[a] | Median | Minimum | Maximum | Medicare rate | % (Max. Min.) ÷ Medicare | Standard deviation | Skewness |
|---|---|---|---|---|---|---|---|---|---|
| 36415 | Collection of venous blood byvenipuncture | $4.36 | $4.29 | $3.18 | $5.64 | $9.09 | 27 | 2.52 | 1.16 |
| 80048 | Basic metabolic panel (BMP) | $7.97 | $7.88 | $5.94 | $10.14 | $8.46 | 50 | 3.56 | 1.44 |
| 80053 | Comprehensive metabolic panel (CMP) | $9.30 | $9.25 | $7.15 | $11.54 | $10.56 | 42 | 4.46 | 1.42 |
| 80061 | Lipid panel | $12.42 | $12.20 | $9.02 | $16.15 | $13.39 | 53 | 6.34 | 1.43 |
| 83036 | Hemoglobin A1c test | $8.55 | $8.48 | $6.70 | $10.49 | $9.71 | 39 | 4.24 | 1.42 |
| 84443 | Thyroid stimulating hormone (TSH) test | $14.97 | $14.97 | $11.21 | $18.76 | $16.80 | 45 | 7.49 | 1.34 |
| 85025 | Complete blood count (CBC) with differential | $7.72 | $7.60 | $6.03 | $9.61 | $7.77 | 46 | 3.15 | 1.47 |
| 85610 | Prothrombin time (PT/INR) | $4.61 | $4.36 | $3.41 | $6.16 | $4.29 | 64 | 1.72 | 1.22 |
| 88305 | Surgical pathology examination, moderate complexity (eg, biopsy tissue exam) | $75.08 | $74.35 | $54.11 | $97.22 | $69.54 | 62 | 41.50 | 1.43 |
| 90471 | Immunization administration (1 vaccine) | $19.42 | $19.38 | $15.04 | $23.87 | $20.05 | 44 | 9.01 | 1.02 |
| 93010 | Electrocardiogram (ECG/EKG) interpretation and report | $10.67 | $10.64 | $9.27 | $12.13 | $7.76 | 37 | 4.19 | 1.09 |
| 97110 | Therapeutic exercises (physical therapy), 15 minutes | $28.19 | $28.05 | $23.59 | $33.00 | $28.79 | 33 | 10.72 | 1.61 |
| 97112 | Neuromuscular reeducation, 15 minutes | $31.37 | $31.18 | $26.36 | $36.62 | $32.02 | 32 | 12.04 | 1.58 |



XSL•FO
RenderX

| Billing code | Billing code description | Mean (averaged over provider TINs)[a] | Median | Minimum | Maximum | Medicare rate | % (Max. Min.) ÷ Medicare | Standard deviation | Skewness |
|---|---|---|---|---|---|---|---|---|---|
| 97140 | Manual therapy techniques (eg, mobilization), 15 minutes | $26.06 | $25.90 | $21.75 | $30.57 | $27.17 | 32 | 10.20 | 1.62 |
| 97530 | Therapeutic activities, 15 minutes | $34.68 | $34.43 | $28.86 | $40.79 | $34.61 | 34 | 13.60 | 1.46 |
| 98941 | Chiropractic manipulation, 3 - 4 spinal regions | $39.45 | $38.50 | $30.83 | $49.27 | $38.49 | 48 | 16.63 | 0.77 |
| 99203 | New patient office visit, moderate complexity | $116.01 | $115.58 | $99.79 | $132.71 | $109.01 | 30 | 36.60 | 1.08 |
| 99212 | Established patient office visit, low complexity | $50.47 | $50.37 | $41.52 | $59.46 | $54.99 | 33 | 17.42 | 0.89 |
| 99213 | Established patient office visit, ow-moderate complexity | $83.27 | $83.02 | $70.93 | $95.93 | $88.95 | 28 | 26.58 | 1.01 |
| 99214 | Established patient office visit, moderate complexity | $121.03 | $120.60 | $103.78 | $138.77 | $125.18 | 28 | 38.73 | 1.03 |

[a]TIN: tax identification number.

Across the 20 billing codes, we find the average variation between the minimum and maximum rates to be 40 % of the Medicare rate (unweighted).

## Discussion

Our assessment shows that providers negotiate a wide range of prices for the same service across insurers. Thus, the same provider can be low cost for one network but high cost for another network. Our findings that prices for commonly utilized services can vary by up to 40% of the Medicare rate in an Office setting compliments existing literature that documents substantial variation in prices across and within hospitals. This highlights the importance of the choice of network and benefit design in reducing health care spending. For employers, health plans, and policymakers, our findings demonstrate that choice of networks and providers can have a substantial impact on health care spending. Future work should aim to understand how navigating patients to lower cost providers can generate potential savings.

### Acknowledgments

The authors acknowledge support for this study from the Peterson Center on Healthcare.

Artificial intelligence was not used in the creation of any portion of this manuscript.

### Funding

Support for this work came from The Peterson Center on Healthcare.



XSL•FO
RenderX

## Data Availability

The analysis in the paper is based on data released by individual insurers under the Transparency in Coverage price transparency rule. These datasets are publicly available on insurers' websites. The authors downloaded, cleaned, and validated the datasets from individual insurers' websites.

## Authors' Contributions

Conceptualization: YP, DM
Data curation: YP, DM
Formal analysis: YP, DM
Funding acquisition: YP, DM
Investigation: YP, DM
Methodology: YP, DM
Project administration: YP, DM
Resources: YP, DM
Software: YP, DM
Supervision: YP, DM
Validation: YP, DM
Visualization: YP, DM
Writing – original draft: YP, DM
Writing – review & editing: YP, DM

## Conflicts of Interest

YP and DM are employed at Simple Healthcare which works with price transparency data.

## References

1. Brot-Goldberg ZC, Chandra A, Handel BR, Kolstad JT. What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics. Q J Econ 2017 Aug 1;132(3):1261-1318. [doi: 10.1093/qje/qjx013]
2. The role of prices in excess US health spending. Health Affairs Brief. 2022 Jun 9. URL: https://www.healthaffairs.org/content/briefs/role-prices-excess-us-health-spending [accessed 2026-07-09]
3. Gul ZG, Sharbaugh DR, Guercio CJ, et al. Large variations in the prices of urologic procedures at academic medical centers 1 year after implementation of the Price Transparency Final Rule. JAMA Netw Open 2023 Jan 3;6(1):e2249581. [doi: 10.1001/jamanetworkopen.2022.49581] [Medline: 36602800]
4. Jiang JX, Forman HP, Gupta S, Bai G. Price variability for common radiology services within U.S. hospitals. Radiology 2023 Mar;306(3):e221815. [doi: 10.1148/radiol.221815] [Medline: 36255313]
5. Whaley C, Radhakrishnan N, Richards M, Simon K, Chartock B. Understanding health care price variation: evidence from Transparency-in-Coverage data. Health Aff Sch 2025 Feb;3(2):qxaf011. [doi: 10.1093/haschl/qxaf011] [Medline: 39916970]
6. Wang Y, Xu J, Anderson G. Within-hospital price gaps across national insurers. JAMA Netw Open 2024 Dec 2;7(12):e2451941. [doi: 10.1001/jamanetworkopen.2024.51941]
7. Muhlestein DB. High prevalence of ghost rates in transparency in coverage data. Health Aff Sch 2025 Nov;3(11):qxaf212. [doi: 10.1093/haschl/qxaf212] [Medline: 41278121]

## Abbreviations

**HCPCS:** Healthcare Common Procedure Coding System
**HPT:** Hospital Price Transparency
**TiC:** Transparency in Coverage
**TIN:** tax identification number

*Edited by M Balcarras; submitted 01.Apr.2026; peer-reviewed by DM Anderson, R Islam; revised version received 17.Jun.2026; accepted 18.Jun.2026; published 23.Jul.2026.*

*Please cite as:*
*Pathak Y, Muhlestein D*
*Within-Provider Variation in Prices for Commonly Utilized Services in Office Setting: Cross-Sectional Study*
*Interact J Med Res 2026;15:e96883*
*URL: https://www.i-jmr.org/2026/1/e96883*
*doi:10.2196/96883*



XSL•FO
RenderX

© Yuvraj Pathak, David Muhlestein. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 23.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

# Effects of a Personalized Retrieval-Augmented Generation Chatbot on Information Needs in Patients With Chronic Kidney Disease: Mixed Methods Randomized Controlled Trial

Hee Jeong Hwang[1], BS; Junghwan Kim[1], BS; Dongseok Heo[2], BS; Jeong Yun You[1], BS; Jeonghwan Lee[3,4], PhD; Soie Kwon[5,6], PhD; Kyung Don Yoo[7], PhD; Jung Han Yoon Park[8], PhD; Bongwon Suh[2,8], PhD; Jung Pyo Lee[3,4*], PhD; Ki Won Lee[1,8,9,10*], PhD

[1]Department of Agricultural Biotechnology, College of Agriculture and Life Sciences, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul, Seoul, Republic of Korea

[10]Institutes of Green Bio Science & Technology, Seoul National University, Pyeongchang, Gangwon State, Republic of Korea

[2]Interdisciplinary Program in Artificial Intelligence, Seoul National University, Seoul, Seoul, Republic of Korea

[3]Department of Internal Medicine, College of Medicine, Seoul National University, Seoul, Seoul, Republic of Korea

[4]Department of Internal Medicine, SNUH SMG-SNU Boramae Medical Center, Seoul, Seoul, Republic of Korea

[5]Department of Internal Medicine, Chung-Ang University Hospital, Seoul, Seoul, Republic of Korea

[6]Department of Internal Medicine, Chung-Ang University, Seoul, Seoul, Republic of Korea

[7]Department of Internal Medicine, Ulsan University Hospital, University of Ulsan College of Medicine, Ulsan, Republic of Korea

[8]Advanced Institutes of Convergence Technology, Seoul National University, Suwon, Gyeonggi-do, Republic of Korea

[9]Center for Food and Bio convergence and Bio-MAX Institute, Seoul National University, Seoul, Seoul, Republic of Korea

[*]these authors contributed equally

**Corresponding Author:**
Ki Won Lee, PhD
Department of Agricultural Biotechnology, College of Agriculture and Life Sciences, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul, Seoul, Republic of Korea

## Abstract

**Background:** Patients receiving dialysis for chronic kidney disease encounter complex self-management challenges and ongoing information needs that traditional health care often fails to meet. Retrieval-augmented generation (RAG) enhances the factual accuracy of large language models (LLMs), but whether personalized RAG is more effective than general RAG remains unclear.

**Objective:** This study aimed to assess whether personalized constraint retrieval-augmented generation (C-RAG) is more effective than open-domain general retrieval-augmented generation (O-RAG) and standard care in reducing information needs among patients receiving dialysis and to identify predictors of chatbot engagement.

**Methods:** A sequential 2-phase mixed methods study was conducted. Phase 1 evaluated 4 LLMs (Claude 3.5 Sonnet [Anthropic], Gemini 1.5 Flash [Google], CLOVA X [NAVER], and ChatGPT-4o [OpenAI]) using 60 patient-generated questions, assessed by 15 clinical experts. Phase 2 was a 3-arm parallel pilot randomized controlled trial (RCT; October-December 2025) at 3 tertiary hospitals in South Korea. Based on phase 1 findings, 2 chatbot variants were developed using GPT-4o—C-RAG, which incorporates patient-specific clinical parameters, and O-RAG, which uses the same knowledge base without personalization. Participants were randomized 1:1:1 to C-RAG, O-RAG, or standard care. The primary outcome, change in medication, dietary, and diagnostic information needs (5-point Likert scales) from baseline to 4 weeks, was analyzed using ANCOVA. Semistructured interviews with 9 participants explored underlying mechanisms.

**Results:** ChatGPT-4o achieved the highest scores in clinical accuracy (mean 3.60, SD 0.95), safety (mean 3.67, SD 1.00), and readability (Flesch-Kincaid Grade Level 9.8, compared with 13.2 - 14.5 for other models). In phase 2, 42/45 (93%) enrolled participants completed the study (14 per group). ANCOVA revealed significant group differences in medication ($F_{2,38}$=4.574, $P$=.02) and dietary ($F_{2,38}$=4.232, $P$=.02) information needs. In Bonferroni-corrected comparisons, C-RAG had lower needs than O-RAG in both domains (observed mean 2.93, SD 1.21 vs mean 3.86, SD 0.95; Cohen $d$=−0.86, 95% CI −1.63 to −0.08; $P$=.02 for both). Against control, C-RAG showed medium-to-large but nonsignificant effects (medication Cohen $d$=−0.64, 95% CI −1.39 to 0.12; dietary Cohen $d$=−0.54, 95% CI −1.30 to 0.21). Secondary outcomes did not differ between groups. Dietary management emerged as the most frequent query topic, accounting for 29.3% (205/691) of total queries. Participants on dialysis for 1 - 3 years showed higher engagement (9/10, 90% engagement for 1 - 3 y vs 5/18, 27.8% for ≥4 y; Fisher exact $P$=.006).



XSL•FO
RenderX

Qualitative analysis revealed 5 themes, including dietary management as the primary need and dialysis duration as a moderator of utility.

**Conclusions:** In this pilot trial, personalized RAG showed lower information needs than general RAG, with a large effect size (Cohen *d*=−0.86). The medium-to-large effects against standard care (Cohen *d*=−0.54 to −0.64) support a fully powered confirmatory trial (estimated 50 - 55 participants per group) to evaluate long-term clinical outcomes.

**Trial Registration:** KDCA CRIS KCT0011656; https://cris.nih.go.kr/cris/search/detailSearch.do?seq=32626

***(Interact J Med Res 2026;15:e96501)*** doi:10.2196/96501

**KEYWORDS**

renal dialysis; chronic kidney disease; chatbots; retrieval-augmented generation; large language models; self-management; dietary management; information needs; digital health; randomized controlled trial

# Introduction

## Background

Chronic kidney disease (CKD) affects over 850 million people globally [1], with a prevalence exceeding 8% in South Korea [2]. Patients with advanced CKD (stages 3 - 5) and those undergoing dialysis face complex self-management challenges, requiring them to simultaneously restrict protein, potassium, phosphorus, sodium, and fluid intake while managing an average of 10 - 12 daily medications [3,4]. This complexity results in ongoing information needs, such as “Can I eat this food?” and “When should I take this medication?“—questions that traditional health care delivery often struggles to address [5,6]. For example, a generic recommendation to “increase fluid and potassium-rich fruit intake for general health” can precipitate hyperkalemia or fluid overload in a patient receiving hemodialysis who is anuric, for whom potassium and fluid must instead be restricted. Consistent with this concern, recent work argues that retrieval-grounded conversational systems require evaluation frameworks more rigorous than those applied to generic chatbot outputs, particularly for risk-sensitive clinical content [7].

Current methods of providing information have significant limitations. Educational pamphlets offer generalized content without customization [6], while online search engines may provide unreliable information lacking CKD-specific context [8]. Outpatient consultations, which typically last 5 - 10 minutes, cannot provide comprehensive patient education [8,9]. Compounding these structural barriers, clinicians often prioritize acute physiological parameters over patient education during these brief encounters [10-12], and they frequently use inconsistent terminology without verifying patient comprehension [13,14]. This care gap forces patients to navigate complex treatment regimens with inadequate guidance, directly affecting treatment adherence and disease outcomes.

The consequences of insufficient information extend beyond patient frustration. In this study, information needs refer to patients’ self-perceived need for additional disease-related information to support self-management decisions. This construct is distinct from objective knowledge deficits; it captures a subjective state of uncertainty and demand for information, which may persist even when measured knowledge is adequate, and which can be assessed directly by self-report. Patients with CKD report significant unmet information needs regarding medication management, dietary restrictions, and complication prevention [15,16]. The heterogeneous nature of CKD progression and the diverse comorbidity profiles demand highly individualized guidance. Identical questions, such as “Can I eat bananas?” require fundamentally different answers based on dialysis modality, potassium levels, and residual kidney function [17]. Generic information can be not only unhelpful but also potentially harmful when applied inappropriately [18].

## Prior Work

### *Large Language Models in Health Care: Promise and Limitations*

Large language models (LLMs) have shown promise in providing 24×7 access to health information through natural language interaction [19,20]. Their ability to process patient queries in conversational language and generate human-like responses positions them as promising tools for addressing the information access gap in health care. However, standard LLMs have significant limitations. They risk hallucination, generating plausible-sounding but factually incorrect information [21], lack transparent reasoning processes, do not provide verifiable citations, and tend to offer generic responses that fail to consider individual patient contexts [22].

Recent evaluations of LLM performance for kidney disease questions showed generally acceptable accuracy in diagnostic and medication domains [23]. However, nutritional guidance—identified as the most common patient concern in social media and dietary-needs studies of patients receiving dialysis—remained problematic across all tested models [16,24]. More fundamentally, existing systems provide population-level information without considering individual patient characteristics that significantly affect appropriate recommendations [17]. This limitation is not just technical; it reflects a conceptual gap—treating personalized medical advice as a generic information retrieval problem.

### *Retrieval-Augmented Generation: Grounding AI in Verified Knowledge*

Retrieval-augmented generation (RAG) addresses the limitations of LLMs by grounding responses in verified medical literature and clinical guidelines, thereby enhancing factual accuracy and trustworthiness [25,26]. Rather than relying solely on the parametric knowledge acquired during training, RAG systems first retrieve relevant documents from curated knowledge bases (eg, Kidney Disease: Improving Global Outcomes [KDIGO]



XSL•FO
RenderX

guidelines and peer-reviewed nephrology literature) and then generate responses based on this retrieved context. This architecture fundamentally reduces the risk of hallucination while enabling transparent citation of information sources.

Preliminary evidence highlights the effectiveness of RAG in delivering health information. A diabetes education study showed that RAG-enhanced LLMs greatly improved the accuracy and comprehensiveness of responses compared with baseline LLM performance [27]. Similarly, in education of patients with prostate cancer, RAG-based chatbots resulted in higher patient satisfaction and better knowledge retention than traditional static educational materials [28]. However, these studies evaluated generic RAG implementations without addressing the imperative for personalization—the need to tailor information to individual patient parameters.

### *The Critical Gap: Empirical Model Selection and Personalization*

Despite the increasing interest in medical AI, model selection in health information systems often occurs without rigorous empirical validation using real patient questions or assessments from domain-specific experts [21]. This represents a critical methodological gap—researchers typically select LLM foundations based on general benchmarks (eg, massive multitask language understanding and reasoning tasks) rather than through systematic evaluation using actual patient-generated queries assessed by clinical domain experts. Such practices risk deploying systems optimized for irrelevant capabilities while underperforming in crucial areas of patient-facing communication, such as readability, safety, and domain-specific accuracy.

Furthermore, while personalized health information systems have shown benefits in managing diabetes and hypertension [29,30], existing approaches typically use rule-based personalization (eg, "IF diabetic THEN restrict sugar") instead of dynamically integrating comprehensive clinical data. These static personalization strategies lack the flexibility needed to address the complex interactions associated with CKD, where recommendations must consider kidney function, electrolyte levels, dialysis adequacy, medication regimens, and comorbidities simultaneously.

No randomized controlled trials (RCTs) have assessed personalized RAG systems for patients with CKD, leaving the comparative effectiveness of personalized vs generic approaches unestablished. This knowledge gap has significant implications; if personalization offers only marginal benefits, resources should prioritize maximizing accessibility. Conversely, if personalization is found to be transformative, generic systems could do more harm than good by offering misleading guidance with a false sense of authority.

### *Distinguishing Constraint RAG: Technical and Theoretical Contributions*

This study introduces the constraint retrieval-augmented generation (C-RAG) system, which offers 2 key innovations that enhance existing approaches. The first innovation is dynamic clinical data integration. Unlike traditional rule-based systems that rely on fixed decision trees, C-RAG dynamically retrieves patients' relevant clinical parameters—such as current laboratory values, active medications, and dialysis modality. This contextual information is incorporated into both the retrieval query and response generation prompt. As a result, the system can provide responses that take multiple interacting factors into account. For example, it can recommend protein intake based on a patient's current albumin level, dialysis adequacy (Kt/V, where K is dialyzer urea clearance, t is dialysis session time, and V is the volume of distribution of urea), and the presence of proteinuria.

The second innovation is constraint-based response generation. C-RAG uses explicit constraint checking, ensuring that responses adhere to patient-specific safety boundaries derived from individual clinical data. In contrast to general RAG systems that deliver immediate, generic responses without considering the patient's context, C-RAG refrains from providing direct answers when critical clinical parameters are absent. For instance, if a patient asks, "Can I eat bananas?" but their serum potassium data are not available, C-RAG will first prompt the patient to provide their recent potassium level before generating dietary guidance. This information-gathering constraint guarantees that responses are tailored to the individual's clinical context, rather than relying on population-level recommendations that may not be suitable for every patient. More broadly, the constraint-checking and clinical-data integration in C-RAG align with an emerging shift toward auditable, compliance-aware clinical AI, in which generated outputs are separated from a deterministic checking layer and made traceable to patient-specific safety boundaries rather than produced unconditionally [31].

For dietary inquiries, the system calculates allowable ranges for potassium, phosphorus, and protein based on each patient's parameters, filtering or annotating recommendations as necessary. This constraint mechanism sets C-RAG apart from retrieval-only approaches that simply present general guidelines without considering individual needs. Theoretically, we hypothesize that personalization reduces information needs through two complementary mechanisms: (1) efficiency gains—patients receive directly applicable guidance without the need for iterative questioning to clarify context; and (2) confidence enhancement—an explicit acknowledgment of individual parameters fosters trust and decreases the desire for further verification. This positions the reduction of information needs not only as a measure of satisfaction but also as an indicator of the quality of decisional support—patients need less additional information when initial responses effectively address their specific situations.

## Study Objectives

This mixed methods RCT addressed three primary research questions:

1. Does C-RAG more effectively reduce information needs compared with open-domain general retrieval-augmented generation (O-RAG) and standard care among patients receiving dialysis?
2. What patient characteristics and usage patterns predict engagement with RAG-based chatbots in the population with CKD?



XSL•FO
RenderX

3. How does personalization awareness shape patient information-seeking behaviors, perceived utility, and satisfaction?

We hypothesized that C-RAG would significantly reduce information needs by providing efficient, tailored responses that eliminate the repetitive questioning cycles often seen with generic information. Additionally, we predicted that awareness of personalization would be a key factor influencing the system's utility, as patients would recognize and appreciate individualized guidance.

The study used a sequential 2-phase design. In phase 1, we systematically evaluated 4 leading LLMs with real patient questions to determine the best foundation model, based on expert assessments of accuracy, safety, and readability. Phase 2 involved a 3-arm RCT that compared C-RAG, O-RAG, and standard care, incorporating qualitative interviews to better understand the mechanisms behind the quantitative outcomes.

# Methods

## Study Design and Overview

We conducted a sequential 2-phase mixed methods study to develop and validate a personalized RAG-based health information system for patients with CKD undergoing dialysis. In phase 1, we used a systematic comparative evaluation framework to identify the optimal LLM foundation by assessing performance in 3 domains—semantic consistency, readability, and clinical validity, using authentic patient-generated questions. Phase 2 involved a 3-arm parallel RCT to assess the clinical effectiveness of C-RAG compared with O-RAG and standard care. This phase also included qualitative interviews to explore the mechanisms behind the quantitative outcomes. This integrated approach facilitated evidence-based progress from technical model selection to clinical implementation while capturing both outcome effectiveness and patient experience.

## Phase 1: Systematic Evaluation and Selection of Foundation Language Model

### *Methodological Framework*

Between October and December 2024, we conducted a comprehensive comparative evaluation of 4 leading LLMs to assess their ability to provide accurate, safe, and accessible responses to inquiries from patients with kidney disease in the Korean language context. The complete evaluation workflow is illustrated in Multimedia Appendix 1.

### *Patient Question Collection and LLM Response Generation*

We used publicly accessible patient-generated content from KidneyCafe, the largest online kidney disease community in South Korea, which has over 15,000 members. All collected questions were fully anonymized. Since the selected content consisted of general educational inquiries posted publicly without identifiable information, institutional review board (IRB) approval was not required. All identifying elements were removed before analysis. Building on our previously published latent Dirichlet allocation topic modeling research [16], we focused on 3 high-priority clinical domains—examination and diagnostics (laboratory interpretation, symptom evaluation, disease monitoring, and costs); medication and dietary supplements (prescription management, over-the-counter drug safety, and supplement appropriateness); and nutrition and diet (addressing food restrictions, electrolyte management, and dietary guidance for various stages of CKD and dialysis modalities). Within each domain, we selected 20 representative questions (total n=60) based on view count rankings, using high community engagement as a proxy for widespread patient interest. This pragmatic sampling prioritized questions that reflect common patient concerns, although it may underrepresent rare or complex cases with lower engagement.

We evaluated 4 prominent LLMs—Claude 3.5 Sonnet (Anthropic), Gemini 1.5 Flash (Google), CLOVA X (NAVER), and ChatGPT-4o (OpenAI). Each model generated 3 independent responses per question through separate chat sessions conducted on October 23 - October 24, 2024. We ensured strict session isolation to prevent context accumulation. Queries were made using default parameters, without fine-tuning or prompt engineering, to preserve the models' out-of-the-box performance. Each model was accessed through its publicly available web interface (Claude [Anthropic], ChatGPT [OpenAI], Gemini [Google], and CLOVA X [NAVER]). This approach was chosen to evaluate the baseline capabilities of the models as experienced by typical end users.

### *Semantic Consistency and Readability*

We used BERTScore (bidirectional encoder representations from transformers score) to quantify semantic consistency across the 3 independent responses generated by each model. For each model's 3 responses per question, we conducted pairwise comparisons, resulting in 3 similarity scores. We established a consistency threshold of 0.7 to identify acceptable semantic alignment, based on prior validation studies showing that BERTScore $F_1$ values above 0.7 indicate substantial semantic similarity [32]. The response with the highest mean BERTScore was selected as the most representative for readability analyses and expert evaluations, focusing on typical outputs while minimizing evaluation burden.

We evaluated text accessibility using 3 complementary readability metrics—the Flesch-Kincaid Grade Level (FKGL), which indicates the years of education required; the Flesch Reading Ease (FRE), which uses a 0 - 100 scale where higher scores indicate easier readability; and the Gunning Fog Index (GFI), which estimates the formal education years needed. These metrics were adapted for Korean using the *KoNLPy* library for morphological analysis. To conduct statistical comparisons, we employed the Kruskal-Wallis H test along with Dunn post hoc pairwise comparisons and Bonferroni correction, analyzing both overall model differences and domain-specific performance patterns. Because Flesch-based metrics were originally developed for English, their direct application to Korean is approximate. We used KoNLPy morphological analysis to define word and syllable boundaries appropriate to Korean agglutinative structure, computing syllable counts per morpheme-segmented word and sentence length from analyzer-defined boundaries. These adapted indices were used



XSL•FO
RenderX

only for relative comparison across models evaluated on identical Korean inputs, not as absolute grade-level claims; this limitation is noted in the Discussion section.

### *Expert Evaluation of Clinical Accuracy and Safety*

We assembled a multidisciplinary panel of 15 board-certified health care professionals, specifically selected to ensure expertise in their respective domains—5 nephrologists, 5 licensed pharmacists, and 5 clinical dietitians. Each evaluator had at least 5 years of specialized clinical experience in managing kidney disease and reported no financial conflicts of interest with AI technology companies.

The expert evaluation used standardized rating protocols to assess the accuracy and safety of responses across 3 clinical domains. For each of the 60 questions, experts reviewed answers from all 4 LLM models as well as community-sourced responses, using two 5-point Likert scales—one for accuracy (1=completely inaccurate, misleading information; 5=completely accurate, comprehensive explanation) and another for safety (1=dangerous, high risk of patient harm; 5=completely safe, appropriate precautions emphasized). Evaluators received standardized instructions outlining the rating criteria, emphasizing the assessment of medical correctness, completeness, appropriateness for the patient audience, and the potential for misinterpretation or harm. To ensure expertise matched to the domain, nephrologists evaluated diagnostic and examination questions, pharmacists evaluated medication and supplement questions, and dietitians evaluated nutrition and diet questions. Each expert independently rated all questions within their domain, resulting in 5 independent ratings per response. Evaluators were blinded to the identity of the models during the assessment, with responses presented in randomized order across evaluation sessions. Community-sourced responses were authentic forum answers and were not length-, context-, or style-matched to the LLM outputs; they were included to represent the real-world peer-information baseline that patients currently encounter. This lack of matching is acknowledged as a limitation when interpreting the community-vs-LLM comparison.

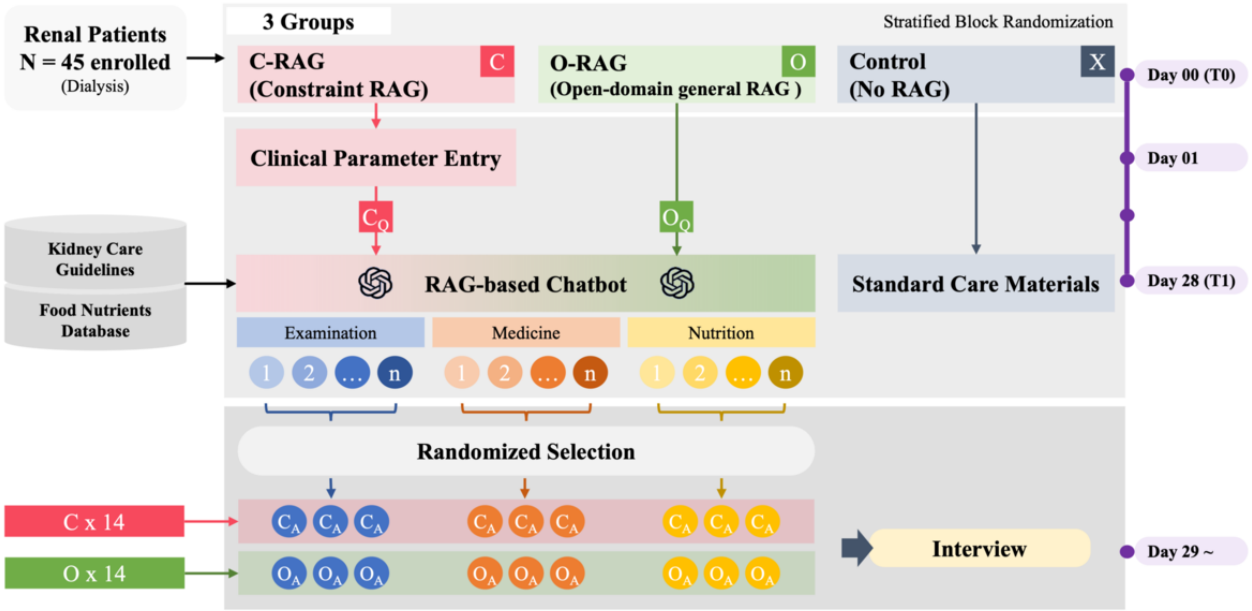
We calculated interrater reliability using Cronbach α to measure expert agreement within each domain. Statistical analysis involved the Kruskal-Wallis H test to evaluate overall differences across the 5 information sources (4 LLMs plus community responses), followed by post hoc pairwise Mann-Whitney U tests with Bonferroni correction when significant omnibus effects were identified. Mean ratings and SDs were computed for each model within each domain and overall, enabling both domain-specific and aggregate performance assessments.

## Phase 2: Clinical Validation of Personalized RAG Through RCT

### *Overview*

Phase 2 addressed the central research question—does C-RAG more effectively reduce patient information needs compared with O-RAG and standard care? We conducted a 3-arm parallel-group pilot RCT, incorporating qualitative interviews to establish preliminary efficacy estimates and explore the mechanisms underlying the quantitative outcomes. The overall study design is illustrated in Figure 1, and the characteristics of the 3 study arms are summarized in Table 1.

**Figure 1.** Phase 2 three-arm parallel randomized controlled trial design. Dialysis patients (n=45 enrolled) were randomly assigned in a 1:1:1 ratio to the C-RAG, O-RAG, or standard care control group. Both RAG groups accessed a smartphone-based chatbot grounded in kidney care guidelines and food nutrient databases over a 4-week intervention period. Assessments were conducted at baseline (T0) and 4 weeks (T1). A subset of intervention participants completed semistructured interviews after the intervention period. C-RAG: constraint retrieval-augmented generation; O-RAG: open-domain general retrieval-augmented generation; RAG: retrieval-augmented generation.






XSL•FO
RenderX

**Table .** Comparison of study arms: chatbot access, knowledge base, and personalization mechanism. Patient-specific clinical data (laboratory values, comorbidities, and medications) were obtained at enrollment and from patient-uploaded documents via OCR[a] rather than direct electronic health record access; refer to the Intervention Development and Implementation section for details.

| Feature | C-RAG[b] | O-RAG[c] | Control |
|---|---|---|---|
| Chatbot access | Yes | Yes | No |
| Knowledge base | Shared curated CKD[d] knowledge base (KDIGO[e], KDOQI[f], KDCA[g], KMA[h], and KSN[i] guidelines) | Shared curated CKD knowledge base (KDIGO, KDOQI, KDCA, KMA, and KSN guidelines) | —[j] |
| Underlying LLM[k] | GPT-4o | GPT-4o | — |
| Patient-specific data integration | Yes (laboratory values, comorbidities, and medications obtained at enrollment and via OCR of patient-uploaded documents) | No | — |
| Constraint checking | Yes (retrieved evidence filtered against patient-specific safety rules before response generation) | No | — |
| Response type | Personalized, patient-bounded recommendations | General educational (population-level) information | Standard care only (no chatbot) |

[a]OCR: optical character recognition.

[b]C-RAG: constraint retrieval-augmented generation.

[c]O-RAG: open-domain general retrieval-augmented generation.

[d]CKD: chronic kidney disease.

[e]KDIGO: Kidney Disease: Improving Global Outcomes.

[f]KDOQI: Kidney Disease Outcomes Quality Initiative.

[g]KDCA: Korea Disease Control and Prevention Agency.

[h]KMA: Korean Medical Association.

[i]KSN: Korean Society of Nephrology.

[j]Not applicable.

[k]LLM: large language model.

### *Study Setting and Participants*

From October to November 2025, we recruited patients receiving dialysis from nephrology outpatient clinics at 3 tertiary medical centers in South Korea—Seoul National University Boramae Medical Center, Ulsan University Hospital, and Chung-Ang University Hospital.

Eligible patients were identified by nephrologists and research nurses during routine outpatient visits at 3 tertiary hospitals (Boramae Hospital, n=26; Ulsan University Hospital, n=9; and Chung-Ang University Hospital, n=10) and approached consecutively. All 45 patients approached were assessed for eligibility, met the inclusion criteria, consented, and were enrolled and randomized; no approached patient declined or was excluded prior to randomization. During follow-up, 3 participants discontinued (C-RAG, n=2, both due to hospitalization for conditions unrelated to the study; O-RAG, n=1, loss of interest; control, n=0), leaving 42 participants (14 per group) for analysis.

Eligibility criteria included age ≥19 years, current maintenance dialysis (hemodialysis or peritoneal dialysis), and having access to a smartphone or computer with basic reading and writing skills suitable for interacting with a text-based system. Basic reading and writing ability was not assessed with a standardized literacy test. Rather, during the consent process, research staff confirmed that each participant could read the study materials and independently complete the self-report questionnaires, which served as the practical criterion for ability to interact with a text-based system. Participants also needed to understand the study procedures and provide written informed consent. Exclusion criteria encompassed cognitive impairments that hindered comprehension of the study content or decision-making ability, limited proficiency in Korean that would restrict questionnaire completion or system use, severe speech or hearing impairments that would prevent participation in interviews, significant comorbidities that would limit participation in the 4-week study, and a history of prior kidney transplantation.

### *Intervention Development and Implementation*

Building on GPT-4o, which was selected in phase 1 for its balance of clinical accuracy and readability, we developed 2 RAG-based chatbot variants—a personalized C-RAG and an O-RAG. This 2-variant design was intended to isolate the contribution of patient-specific personalization beyond RAG retrieval alone. This system used a hybrid retrieval pipeline that combined dense embeddings with keyword searches to extract relevant passages from a curated knowledge base. Text embeddings were created using OpenAI embedding models, and documents were indexed in a vector database that supports semantic similarity searches. Retrieved passages were then



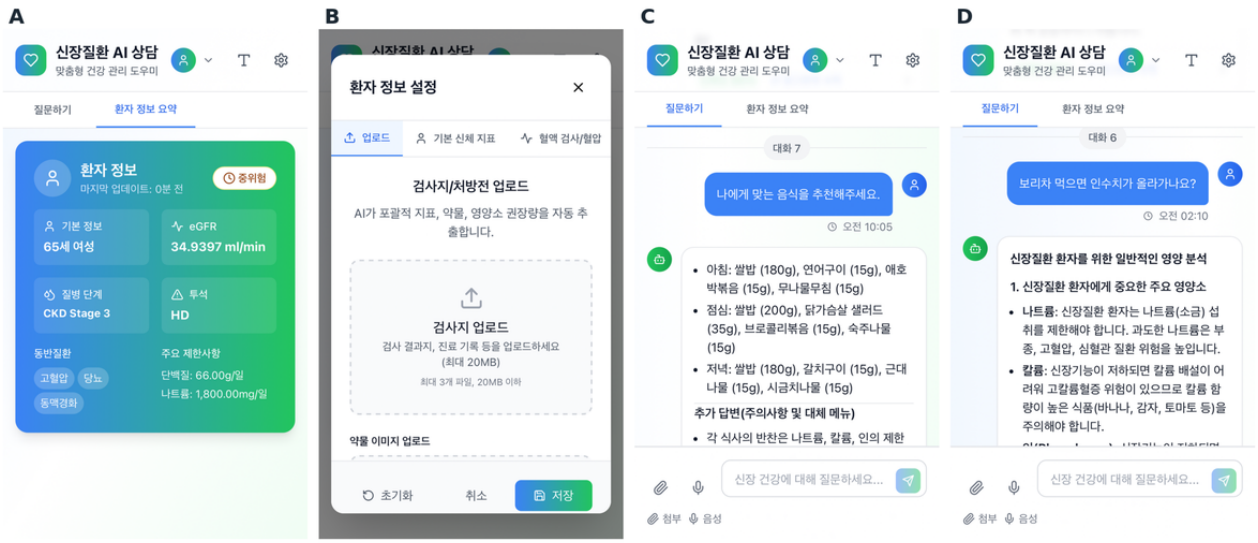
XSL•FO
RenderX

reranked based on their relevance to the query before synthesizing a response.

The knowledge base included well-established clinical and dietary resources, such as multiple KDIGO clinical practice guidelines (2009 - 2024) that cover CKD evaluation, anemia, mineral and bone disorders, diabetes management, blood pressure control, and hepatitis C. It also featured the KDOQI 2020 nutrition guidelines and Korean-language resources from the Korea Disease Control and Prevention Agency, the Korean Medical Association, and the Korean Society of Nephrology. These documents were carefully curated and organized to facilitate the retrieval of evidence pertinent to patient inquiries.

These 2 variants shared the same GPT-4o backbone, retrieval pipeline, and knowledge base, differing only in whether patient-specific constraints were applied. This design allowed attribution of any differential outcome to the personalization mechanism rather than to differences in the underlying model or knowledge source. The C-RAG used patient-specific clinical data—recent laboratory values (eg, serum potassium and phosphorus), comorbidities, and medication information. These data were not drawn from a direct electronic health record interface; instead, they were obtained at enrollment and from patient-uploaded laboratory reports and prescriptions, from which clinical parameters were extracted via optical character recognition (Figure 2B), then used to filter and contextualize responses. Evidence retrieved was subjected to predefined safety rules, and the language model was directed to generate recommendations that adhered to these constraints. To illustrate how the constraint mechanism altered the final answer, consider the query “Can I eat bananas?.” O-RAG returned a general guideline statement that bananas are high in potassium and should be limited in kidney disease. C-RAG, when the patient’s recent serum potassium was unavailable, first requested the value; once a potassium of 5.8 mEq/L was supplied, it advised avoiding bananas and proposed lower-potassium alternatives with portion sizes. A second example involves protein; for “How much meat can I eat?,” O-RAG provided a generic CKD protein range, whereas C-RAG computed a target from the patient’s dialysis modality and recent albumin, yielding a gram-level daily recommendation. Thus, the same retrieved evidence produced population-level guidance in O-RAG but patient-specific, bounded guidance in C-RAG. In contrast, the O-RAG used the same knowledge base and conversational interface but did not incorporate patient-specific constraints, instead offering general educational responses.

**Figure 2.** Screenshots of the constraint retrieval-augmented generation (C-RAG) and open-domain general retrieval-augmented generation (O-RAG) chatbot interfaces. (A) The C-RAG patient summary dashboard displaying automatically populated clinical parameters, including estimated glomerular filtration rate, dialysis modality, comorbidities, and personalized dietary restrictions. (B) The patient information settings panel, through which users upload laboratory reports or prescriptions for automated clinical parameter extraction via optical character recognition. (C) A C-RAG response to a dietary query (“please recommend foods suitable for me”), showing a personalized meal plan with gram-level specifications derived from the patient’s individual clinical parameters. (D) An O-RAG response to a dietary query (“does barley tea raise phosphorus levels?"), illustrating a general population-level nutritional response provided without patient-specific contextualization.




The system interface was delivered through a mobile app optimized for older adults, featuring simplified navigation and large text. Participants could submit free-text questions at any time. Responses were generated through a retrieval process followed by language model synthesis, with constraint checks applied before delivery in the C-RAG condition. Representative screenshots of both system interfaces are presented in Figure 2.

Control participants did not have access to the chatbot and continued with standard dialysis care, which included routine nephrologist visits, dietitian consultations, nursing education during dialysis sessions, and patient education materials. This pragmatic control reflects the usual real-world care against which digital health interventions must demonstrate incremental value.

### *Intervention Delivery*

Following the baseline assessment and randomization, participants in the C-RAG and O-RAG groups received standardized 30-minute individual training sessions. These sessions covered system access procedures, basic navigation,



XSL•FO
RenderX

sample question formulation, and contact information for technical support. The training emphasized the chatbot's educational purpose and encouraged participants to use it for daily decision-making questions (meal planning, medication concerns, and symptom interpretation). Participants were encouraged to ask at least 3 questions weekly across diagnostic, medication, and dietary domains. This threshold was a pragmatic minimum chosen to ensure at least 1 query per domain per week and thereby meaningful exposure to the system; it was not derived from a formal dose-response rationale. No penalties were applied for not meeting it, and adherence to the threshold was not used as an inclusion or analysis criterion. System access remained available 24×7 throughout the intervention period, with usage patterns tracked through automated logging (timestamp, query text, and response content).

### *Randomization and Allocation Concealment*

Randomization sequences were generated by an independent research staff member (not an author) who was not involved in participant recruitment, using computer-generated random numbers with variable block sizes (3 and 6) to minimize predictability in allocation. Allocation concealment was upheld through a centralized web-based registration system, which automatically assigned each participant to a group only after completing enrollment and baseline assessment. This approach ensured that recruitment personnel remained unaware of upcoming assignments. Due to the nature of the intervention, complete blinding was not feasible. Participants were informed of their allocation to either the chatbot group or the control group; however, those in the chatbot groups were not told which specific system they were assigned to (C-RAG vs O-RAG). Research staff responsible for system orientation were unblinded to provide appropriate instruction. Qualitative interview coders remained blinded to group assignments to minimize interpretation bias.

### *Outcome Measures*

All outcomes were assessed at baseline (T0) and 4 weeks post randomization (T1), except for the System Usability Scale (SUS), information accuracy, trustworthiness, and clinical utility, which were evaluated at T1 in the intervention groups only. The 4-week time frame provided a balance between meaningful engagement through intervention exposure and the feasibility of participant retention, minimizing burden in this pilot trial.

The primary end point was the reduction in information needs, assessed across 3 clinically relevant domains—medication, dietary, and diagnostic information. Each domain was measured using a single-item 5-point Likert scale (1=no information needed; 5=urgent information needed). The medication information needs following item asked: "How much do you need information about your kidney disease medications (purpose, dosage, side effects, drug interactions)?" The dietary information needs following item asked: "How much do you need information about appropriate foods and eating patterns for your kidney disease?" The diagnostic information needs following item asked: "How much information do you need to understand your laboratory test results and symptoms?" Owing to the absence of a validated instrument for measuring domain-specific information needs in patients with kidney disease [33], single-item Likert scales were used for each domain, consistent with the approach used in prior chatbot RCTs [28].

For secondary outcomes, disease knowledge was assessed using the Kidney Knowledge Survey (KiKS) [15], a validated 28-item instrument that covers CKD-specific knowledge, including kidney function, blood pressure targets, treatment options, symptoms of disease progression, and medications to avoid. Items are scored dichotomously, and total scores are expressed as the number of correct responses (range 0‐28), with higher scores indicating greater disease knowledge. Self-efficacy was measured using the Chronic Kidney Disease Self-Efficacy scale [34], a 25-item instrument rated on a 5-point Likert scale (1=not at all confident; 5=very confident). Total scores range from 25 to 125, with higher scores indicating stronger self-efficacy. Treatment adherence was assessed across 3 domains—medication adherence, dietary adherence, and fluid restriction adherence. Medication adherence was evaluated using the 8-item Morisky Medication Adherence Scale (MMAS-8) [35], which comprises 7 yes or no items and one 5-point item (total score 0‐8; higher scores indicate better adherence). Dietary adherence and fluid restriction adherence were each assessed using a single 5-point Likert item (1=not at all adherent; 5=very adherent).

In the intervention groups only, system usability was evaluated at T1 using the SUS, a validated 10-item instrument that yields scores from 0 to 100. Scores of 68 or above indicate acceptable usability [36]. Additionally, three 5-point Likert items assessed: (1) satisfaction with the preparation and quality of clinical consultations, (2) satisfaction with the quality of health information provided by the chatbot, and (3) trust in the health information delivered by the chatbot. Usage patterns were captured through automated system logs that recorded query frequency, temporal distribution (day of the week, hour of the day), and topic classification (diagnostic, medication, dietary, lifestyle, and general).

### *Qualitative Interview Substudy*

To enhance the understanding of quantitative outcomes and the mechanisms behind observed effects, we conducted semistructured interviews with a purposive sample of 9 participants (5 from C-RAG and 4 from O-RAG) chosen to maximize variation in engagement levels. This pragmatic sample size was influenced by resource constraints from the pilot trial, while still ensuring representation across both intervention arms and usage patterns.

The interviews were held one-on-one by 2 research team members (HJH and JYY) using a semistructured guide. This guide focused on overall system experience and perceived usefulness, specific examples of information-seeking behaviors, the impact of responses on decision-making, comparisons with alternative information sources (clinicians, internet, and peer communities), and suggestions for improvement. The full semistructured interview guide is provided in Multimedia Appendix 2.



XSL•FO
RenderX

With participant consent, the interviews were audio-recorded and transcribed verbatim by team members. Thematic analysis was conducted following Braun and Clarke's 6-phase framework—familiarization, initial coding, theme development, theme review, theme refinement, and report production [37]. Given the small purposive sample tied to the pilot design, we did not conduct a formal data-saturation assessment; coding continued until the 2 coders (HJH and JYY) reached consensus, and the limited sample is acknowledged as a constraint on thematic completeness. The 2 researchers (HJH and JYY) independently coded the transcripts, resolved discrepancies through discussion, and identified themes through an iterative comparison process.

### Statistical Analysis

#### *Sample Size and Power Considerations*

This study was designed as a pilot RCT, with the primary objective of estimating effect size to inform a future fully powered trial rather than conducting definitive hypothesis testing [38]. Accordingly, between-group comparisons—except the better-powered C-RAG vs O-RAG contrast—are interpreted as effect-size estimates to inform a future confirmatory trial, rather than as definitive tests of efficacy. A target enrollment of 20 participants per group (totaling 60) was established based on G*Power 3.1 (Heinrich Heine University Düsseldorf) calculations for 1-way ANCOVA with 3 groups (80% power, $\alpha$=.05, 1 covariate, Cohen $f$=0.35, pre-post correlation of 0.65), accounting for a 15% anticipated attrition rate. Given the pilot design, the findings should be interpreted as preliminary estimates rather than confirmatory evidence.

#### *Analytic Approach*

Group differences at 4 weeks were assessed using 1-way ANCOVA, incorporating the corresponding baseline score as a covariate. When the omnibus test was significant, pairwise comparisons were conducted using contrast $t$ tests within the same ANCOVA model to ensure consistent baseline adjustment, applying Bonferroni correction for 3 comparisons. Effect sizes are reported as Cohen $d$, interpreted as small (|Cohen $d$|≥0.2), medium (|Cohen $d$|≥0.5), large (|Cohen $d$|≥0.8), and very large (|Cohen $d$|≥1.2). Missing data were addressed through complete-case analysis. Usage patterns were summarized descriptively. All analyses were performed in Python (version 3.11.5; Python Software Foundation) with a 2-tailed $\alpha$=.05.

### Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki. The protocol received IRB approval at all 3 clinical sites (IRB 10-2025-62, IRB 2510-003-660, and UUH IRB 2025-09-025) and at Seoul National University for data analysis (IRB 2508/002 - 001). All participants provided written informed consent, which included explicit notification that their medical care would not be affected if they chose to withdraw at any time. The capacity to provide informed consent was verified prior to enrollment. The phase 1 online community data were exempt from IRB approval, as no individually identifiable information was used. The phase 2 study was registered in the Clinical Research Information Service (registration KCT0011656).

## Results

### Phase 1: LLM Selection Through Systematic Evaluation (ChatGPT-4o)

#### *Semantic Consistency and Readability*

All 4 evaluated LLMs demonstrated acceptable semantic consistency, with BERTScore values exceeding the 0.7 threshold across the 3 independent responses generated per question. ChatGPT-4o achieved superior readability in all 3 metrics (FKGL 9.8, FRE 62.4, and GFI 11.2), indicating a high-school reading level, compared with Claude, Gemini, and CLOVA X (FKGL 13.2 - 14.5, college-level). The Kruskal-Wallis H test confirmed significant differences in readability across the models ($P$<.001; Multimedia Appendix 3 ).

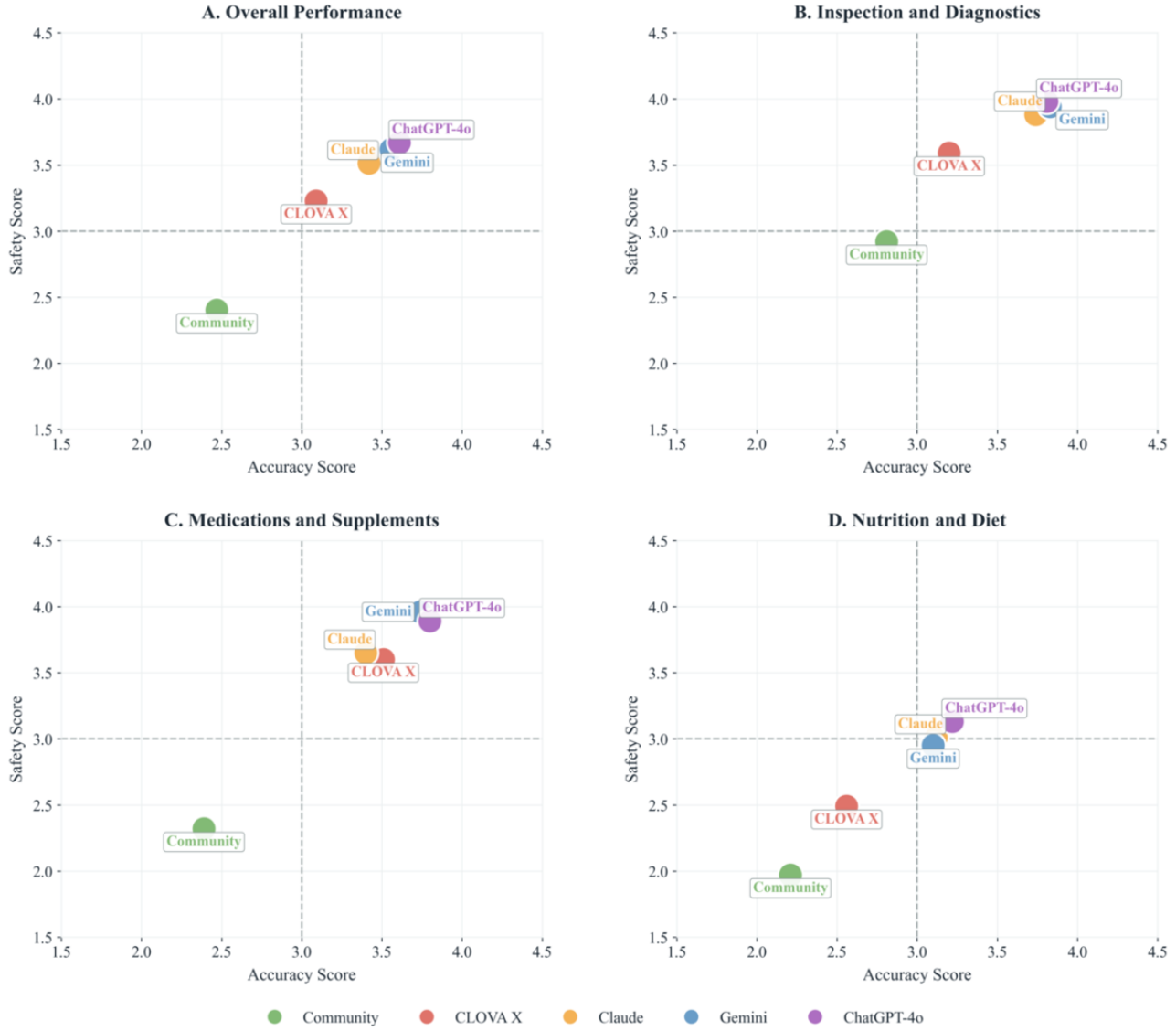
#### *Expert Evaluation of Clinical Accuracy and Safety*

Interrater reliability was strong across all evaluations (Cronbach $\alpha$ 0.802 - 0.984). ChatGPT-4o achieved the highest clinical accuracy (mean 3.60, SD 0.95) and safety ratings (mean 3.67, SD 1.00), followed by Gemini, which had an accuracy mean of 3.54 (SD 1.07) and a safety mean of 3.62 (SD 1.08). There was no significant difference between ChatGPT-4o and Gemini (Mann-Whitney $U$ test). Claude demonstrated intermediate performance with an accuracy mean of 3.32 (SD 1.12) and a safety mean of 3.45 (SD 1.15), while CLOVA X showed lower accuracy (mean 3.08, SD 1.18). All 4 LLMs substantially outperformed community-sourced responses from KidneyCafe patient forums, which had an accuracy mean of 2.49 (SD 1.21) and a safety mean of 2.44 (SD 1.28; Kruskal-Wallis H test, $P$<.001). Domain-specific analysis revealed that all models performed comparably in the inspection or diagnostics and medications or supplements domains, while the nutrition and diet domain yielded lower scores across all models (Figure 3).



XSL•FO
RenderX

**Figure 3.** Clinical accuracy and safety scores across large language models (LLMs) by domain. Each panel displays the mean accuracy (x-axis) and safety (y-axis) scores for 4 LLMs and community-sourced responses. Dashed lines indicate score 3.0 (acceptable threshold). ChatGPT-4o and Gemini consistently occupied the high-accuracy, high-safety quadrant across all domains. The nutrition and diet domain exhibited the lowest overall performance among all models, including ChatGPT-4o. Community-sourced responses consistently fell below the acceptable threshold in both accuracy and safety dimensions.




### *Model Selection*

ChatGPT-4o's combination of top-tier clinical validity (accuracy mean 3.60, SD 0.95; safety mean 3.67, SD 1.00), superior patient accessibility (FKGL 9.8 vs 13‐14 for competitors), and consistent cross-domain performance established it as the optimal foundation model. While Gemini demonstrated comparable clinical scores, ChatGPT-4o's 4-grade-level readability advantage makes it more appropriate for diverse patient populations with varying educational backgrounds. ChatGPT-4o was therefore selected for phase 2 RAG system development.

## Phase 2: Clinical Validation of Personalized RAG Through RCT

### *Participant Flow and Baseline Characteristics*

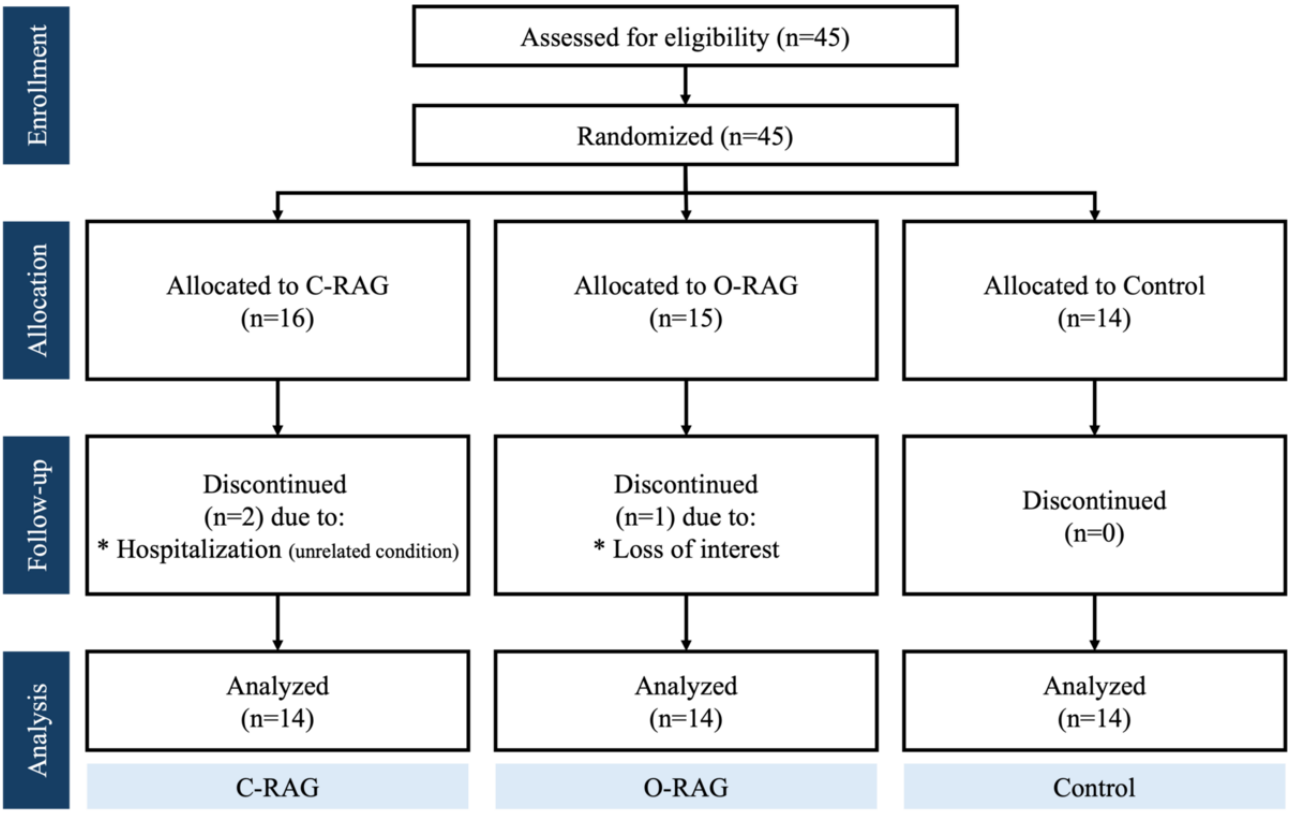
Between October and November 2025, 45 dialysis patients were evaluated for eligibility across 3 hospital sites. After randomization, 3 participants withdrew from the intervention period—2 from the C-RAG due to hospitalization for unrelated conditions and one from the O-RAG due to loss of interest—resulting in 42 completers (14 per group; 93% retention rate).

The incidence of withdrawal was consistently low across all study groups (C-RAG 2/16, O-RAG 1/15, and control 0/14) and was unrelated to the intervention, supporting a complete-case analysis. Although enrollment fell short of the planned 60 participants due to recruitment constraints at various sites during the study period, the achieved sample aligns with precedents from pilot RCTs in CKD digital health interventions [27,28] and is sufficient for effect size estimation. All 42 completers provided usable data for analysis (Figure 4).



XSL•FO
RenderX

**Figure 4.** CONSORT (Consolidated Standards of Reporting Trials) flow diagram of participant enrollment, randomization, and analysis. C-RAG: constraint retrieval-augmented generation; O-RAG: open-domain general retrieval-augmented generation.




Baseline characteristics revealed no statistically significant differences among the groups (Table 2). Age distribution was comparable ($P$=.70), with the majority of participants in their 40s-50s across all groups. Sex distribution showed male predominance (12/14, 85.7% C-RAG; 12/14, 85.7% O-RAG; and 11/14, 78.6% control; $P$=.84). Dialysis modality was predominantly hemodialysis across all groups (13/14, 92.9% C-RAG; 14/14, 100% O-RAG; and 14/14, 100% control), with 1 C-RAG participant on peritoneal dialysis ($P$≥.99). Dialysis duration distribution was comparable across groups ($P$=.50), with the majority of participants having ≥5 years of dialysis experience. Comorbidity prevalence was comparable across groups—diabetes mellitus ($P$=.08), hypertension ($P$=.10), and cardiovascular disease ($P$=.86). Baseline kidney knowledge, self-efficacy, information needs, and adherence measures showed no significant differences across groups ($P$=.07 - .82).



XSL•FO
RenderX

**Table .** Baseline characteristics of participants by group.

| Characteristic | C-RAG[a] (n=14) | O-RAG[b] (n=14) | Control (n=14) | *P* value[c] |
|---|---|---|---|---|
| Age (y), n (%) | | | | .70 |
| 40s | 4 (28.6) | 2 (14.3) | 4 (28.6) | |
| 50s | 4 (28.6) | 8 (57.1) | 4 (28.6) | |
| 60s | 2 (14.3) | 0 (0) | 2 (14.3) | |
| 70s | 2 (14.3) | 3 (21.4) | 3 (21.4) | |
| 80+ | 2 (14.3) | 1 (7.1) | 1 (7.1) | |
| Sex, n (%) | | | | .84 |
| Male | 12 (85.7) | 12 (85.7) | 11 (78.6) | |
| Female | 2 (14.3) | 2 (14.3) | 3 (21.4) | |
| Dialysis modality, n (%) | | | | ≥.99 |
| Hemodialysis | 13 (92.9) | 14 (100) | 14 (100) | |
| Peritoneal dialysis | 1 (7.1) | 0 (0) | 0 (0) | |
| Dialysis duration, n (%) | | | | .50 |
| <6 months | 2 (14.3) | 0 (0) | 0 (0) | |
| 6 months-1 year | 1 (7.1) | 2 (14.3) | 1 (7.1) | |
| 1-3 years | 2 (14.3) | 3 (21.4) | 1 (7.1) | |
| 3-5 years | 4 (28.6) | 2 (14.3) | 3 (21.4) | |
| ≥5 years | 5 (35.7) | 7 (50) | 9 (64.3) | |
| Comorbidities, n (%) | | | | |
| Diabetes mellitus | 10 (71.4) | 4 (28.6) | 7 (50) | .08 |
| Hypertension | 10 (71.4) | 13 (92.9) | 8 (57.1) | .10 |
| Cardiovascular disease | 2 (14.3) | 3 (21.4) | 3 (21.4) | .86 |
| Baseline measures, mean (SD) | | | | |
| Kidney knowledge (KiKS[d]) | 18.6 (3.4) | 19.6 (3.1) | 19.0 (3.7) | .82 |
| Self-efficacy (CKD-SE[e]) | 81.4 (12.9) | 86.7 (14.1) | 86.8 (16.2) | .58 |
| Diagnosis info needs | 3.2 (1.1) | 3.2 (1.0) | 3.4 (1.2) | .81 |
| Medication info needs | 3.8 (0.8) | 2.9 (1.1) | 3.6 (1.2) | .07 |
| Diet info needs | 3.6 (1.2) | 3.3 (1.1) | 3.9 (1.1) | .38 |
| Medication adherence (MMAS-8[f]) | 6.2 (1.4) | 5.7 (2.0) | 6.0 (2.0) | .69 |
| Diet adherence | 2.6 (0.7) | 3.4 (1.0) | 3.1 (0.7) | .08 |
| Fluid adherence | 2.9 (0.8) | 3.4 (1.0) | 2.9 (0.5) | .07 |

[a]C-RAG: constraint retrieval-augmented generation.

[b]O-RAG: open-domain general retrieval-augmented generation.

[c]*P* values were calculated using the Kruskal-Wallis test for continuous variables and the chi-square test or Fisher exact test for categorical variables.

[d]KiKS: Kidney Knowledge Survey.

[e]CKD-SE: Chronic Kidney Disease Self-Efficacy Scale.

[f]MMAS-8: Morisky Medication Adherence Scale-8.

### *Primary Outcomes: Information Needs*

A 1-way ANCOVA adjusting for baseline scores revealed significant group differences in medication information needs ($F_{2,38}$=4.574, *P*=.02) and dietary information needs ($F_{2,38}$=4.232, *P*=.02) at 4 weeks. No significant group difference was found for diagnostic information needs ($F_{2,38}$=0.835, *P*=.44; Table 3).



XSL•FO
RenderX

**Table .** Primary outcomes: information needs at 4 weeks. Values are observed means at 4 weeks (SD) on a 1 - 5 Likert scale; lower scores indicate lower information needs.

| Domain and comparison | C-RAG[a], mean (SD) | O-RAG[b], mean (SD) | Control, mean (SD) | Effect size, Cohen *d*[c] (95% CI) | *P* value |
|---|---|---|---|---|---|
| Medication | 2.93 (1.21) | 3.86 (0.95) | 3.71 (1.27) | | |
| ANCOVA[d] | | | | —[e] | .02 |
| C-RAG vs O-RAG | | | | −0.86 (−1.63 to −0.08) | .02 |
| C-RAG vs Control | | | | −0.64 (−1.39 to 0.12) | .13 |
| O-RAG vs Control | | | | 0.13 (−0.61 to 0.87) | ≥.99 |
| Dietary | 2.93 (1.21) | 3.86 (0.95) | 3.57 (1.16) | | |
| ANCOVA[d] | | | | — | .02 |
| C-RAG vs O-RAG | | | | −0.86 (−1.63 to −0.08) | .02 |
| C-RAG vs Control | | | | −0.54 (−1.30 to 0.21) | .55 |
| O-RAG vs Control | | | | 0.27 (−0.47 to 1.01) | .40 |
| Diagnostic | 3.43 (1.22) | 3.93 (0.83) | 3.71 (1.07) | | |
| ANCOVA[d] | | | | — | .44 |
| C-RAG vs O-RAG | | | | −0.48 (−1.23 to 0.27) | .61 |
| C-RAG vs Control | | | | −0.25 (−0.99 to 0.50) | ≥.99 |
| O-RAG vs Control | | | | 0.22 (−0.52 to 0.97) | ≥.99 |

[a]C-RAG: constraint retrieval-augmented generation.

[b]O-RAG: open-domain general retrieval-augmented generation.

[c]Cohen *d* calculated from raw observed group values.

[d]ANCOVA omnibus test ($F_{2,38}$) adjusted for corresponding baseline score. Pairwise comparisons via contrast *t* tests within ANCOVA model with Bonferroni correction for 3 comparisons. Medication and dietary information-needs scores were highly correlated within participants ($r$=0.49 - 0.85 across groups, including at baseline), reflecting a general tendency for patients to rate these closely related domains similarly; the resulting identical group-level means, SDs, and effect sizes are a consequence of this correlation pattern rather than a data-entry artifact.

[e]Not applicable.

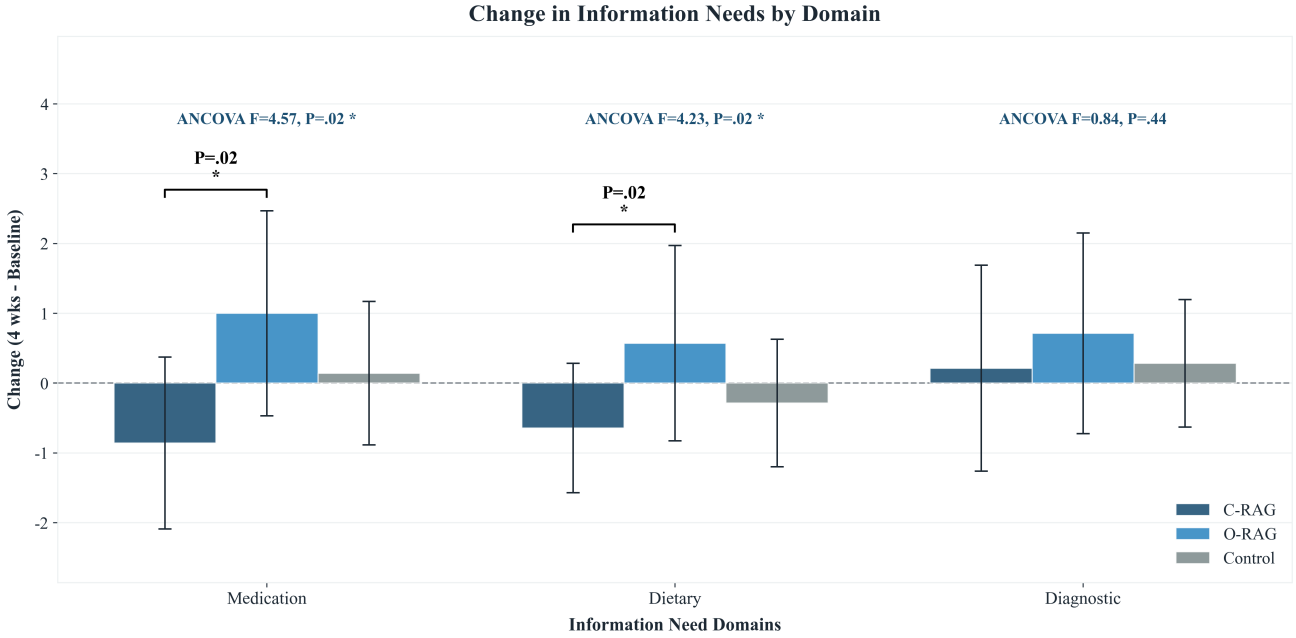
Pairwise comparisons using Bonferroni-corrected contrast *t* tests within the ANCOVA model indicated that C-RAG participants had significantly lower information needs than O-RAG participants at 4 weeks for both the medication (Cohen *d*=−0.86, 95% CI −1.63 to −0.08; *P*=.02) and dietary domains (Cohen *d*=−0.86, 95% CI −1.63 to −0.08; *P*=.02), representing large effect sizes. No significant differences were found between C-RAG and the control group, or between O-RAG and the control group, for any domain (medication Cohen *d*=−0.64, 95% CI −1.39 to 0.12; dietary Cohen *d*=−0.54, 95% CI −1.30 to 0.21).

Examination of change scores revealed divergent trajectories between the two RAG groups (Figure 5). C-RAG participants demonstrated reductions in information needs from baseline in both the medication (Δ=−0.87) and dietary (Δ=−0.67) domains, indicating improved fulfillment of information needs. In contrast, O-RAG participants experienced increases in information needs in both domains (Δ=+0.96 and Δ=+0.56, respectively), although these increases did not differ significantly from the control group. This pattern suggests a potential trend toward heightened awareness of unmet information needs, although the lack of statistical significance limits definitive interpretation. Control participants exhibited minimal change across all domains.



XSL•FO
RenderX

**Figure 5.** Information needs change scores by domain and group. Constraint retrieval-augmented generation (C-RAG) demonstrated large effect sizes for information needs reduction vs open-domain general retrieval-augmented generation (O-RAG) in medication (Cohen *d*=−0.86, *P*=.02) and dietary domains (Cohen *d*=−0.86, *P*=.02). Error bars represent standard deviation. **P*<.05, trend (*P*<.10).




### *Usability and Satisfaction*

In the intervention groups, the mean SUS score was 62.50 (SD 15.60) for C-RAG and 66.96 (SD 14.91) for O-RAG, both marginally below the 68-acceptability threshold, with no significant between-group difference (*P*=.55). Satisfaction with consultation preparation or quality (C-RAG mean 3.50, SD 0.65; O-RAG mean 3.79, SD 1.12; *P*=.24), chatbot information quality (mean 3.64, SD 0.74 vs mean 3.79, SD 0.70; *P*=.74), and trust in chatbot information (mean 3.57, SD 0.65 vs mean 3.86, SD 0.66; *P*=.35) likewise did not differ between groups.

### *Secondary Outcomes*

Secondary outcomes including kidney disease knowledge (KiKS), self-efficacy (Chronic Kidney Disease Self-Efficacy scale), and treatment adherence showed no significant group differences over the 4-week intervention period (Multimedia Appendix 4), consistent with the short duration of this pilot study.

### *Chatbot Usage Patterns*

System logs provided comprehensive usage data from 28 intervention participants (14 C-RAG and 14 O-RAG) throughout the 4-week study period, offering objective behavioral evidence of how patients integrated the chatbot into their daily disease management routines.

Participants generated 691 questions collectively, yielding a mean of 24.68 (SD 24.29) questions per participant (median 13.0, IQR 4.0 - 40.0). Usage frequency demonstrated marked heterogeneity, with participants clustering into 2 distinct engagement profiles. Because no validated cutoff for chatbot engagement exists, participants were dichotomized at the sample median (13.0 questions) into higher- and lower-engagement groups for exploratory within-sample comparison only. Higher-engagement users (n=14) averaged 44.8 (SD 18.5) questions, whereas lower-engagement users (n=14) averaged 4.6 questions. C-RAG and O-RAG groups showed comparable usage levels (mean 20.00 (SD 20.15) vs mean 29.36 (SD 27.79) questions; *P*=.40), indicating that personalization affected information quality rather than engagement quantity.

To identify patient populations most likely to benefit from chatbot interventions, we conducted an exploratory analysis comparing higher- and lower-engagement users (Table 4). Dialysis duration emerged as the only significant predictor (*P*=.006)—9 of 10 (90%) patients with shorter dialysis duration (1 - 3 y) showed high engagement vs 5 of 18 (27.8%) with longer duration (≥4 y). Sex showed no significant association with engagement (*P*=.11). Age, dialysis modality, diabetes status, digital device proficiency, and prior chatbot experience were not associated with engagement (Table 4), suggesting broad accessibility regardless of technical background.



XSL•FO
RenderX

**Table .** Predictors of high chatbot engagement.

| Characteristic | High engagement[a] (n=14) | Low engagement (n=14) | *P* value[b] |
|---|---|---|---|
| Sex, n (%) | | | .11 |
| Female | 4 (100) | 0 (0) | |
| Male | 10 (41.7) | 14 (58.3) | |
| Dialysis duration (y), n (%) | | | .006 |
| 1-3 | 9 (90) | 1 (10) | |
| ≥4 | 5 (27.8) | 13 (72.2) | |
| Dialysis modality, n (%) | | | ≥.99 |
| HD[c] | 14 (51.9) | 13 (48.1) | |
| PD[d] | 0 (0) | 1 (100) | |
| Diabetes mellitus, n (%) | | | .71 |
| Yes | 6 (42.9) | 8 (57.1) | |
| No | 8 (57.1) | 6 (42.9) | |
| Digital device proficiency, mean (SD) | 3.4 (1.0) | 3.3 (0.8) | .64 |
| Prior chatbot experience, n (%) | | | .45 |
| Yes | 8 (61.5) | 5 (38.5) | |
| No | 6 (40) | 9 (60) | |

[a]High engagement defined as above the median number of questions submitted during the 4-week intervention period (median 13.0, range 0‐89). Analysis restricted to C-RAG and O-RAG groups (n=28).

[b]*P* values were calculated using the Fisher exact test for categorical variables and the Mann-Whitney *U* test for continuous variables.

[c]HD: hemodialysis.

[d]PD: peritoneal dialysis.

### *Information-Seeking Priorities and Temporal Integration*

Topic distribution analysis (Multimedia Appendix 5A) showed that dietary queries comprised the largest single category (205/691, 29.3%), followed by diagnostic (156/691, 22.6%) and medication (122/691, 17.7%) queries; keyword-level frequencies are provided in Multimedia Appendix 5B.

Temporal usage revealed distinct patterns; engagement peaked in week 1 (276 questions) and tapered toward week 4 (107 questions), and was higher on weekdays than weekends, peaking Tuesday through Thursday (Multimedia Appendix 5C). Hourly analysis identified morning (9 AM) and evening (9 PM) peaks that were dietary-dominant, whereas midday peaks (2 PM‐3 PM) showed a more varied mix of diagnostic and lifestyle queries (Multimedia Appendix 5D).

## Qualitative Findings: Personalization Awareness Fundamentally Transforms Patient Experience

### *Overview*

A total of 9 patients receiving dialysis completed semistructured interviews at the study end point—5 from the C-RAG group and 4 from the O-RAG group. Table 5 presents the participant characteristics, including dialysis duration and levels of system engagement. Moreover, 5 major themes emerged from the analysis, providing mechanistic insight into quantitative outcome patterns.



XSL•FO
RenderX

**Table .** Characteristics of qualitative interview participants. System usage categories based on total number of sessions during the 4-week intervention period: minimal (1-5), low (6-15), moderate (16-50), high (>50)

| ID | Group | Age group (y) | Sex | Dialysis duration | Diagnosis duration | System usage and total session count |
|---|---|---|---|---|---|---|
| O1 | O-RAG[a] | 50 - 59 | Female | ≥5 years | ≥10 years | High (54) |
| C3 | C-RAG[b] | 30 - 49 | Male | ≥5 years | ≥10 years | Low (7) |
| O4 | O-RAG | 65 - 74 | Male | 6 months-1 year | ≥10 years | High (89) |
| C5 | C-RAG | 65 - 74 | Male | ≥5 years | 5 - 10 years | Moderate (45) |
| C16 | C-RAG | 50 - 59 | Female | 6 months-1 year | 1 - 3 years | Moderate (33) |
| O19 | O-RAG | 30 - 49 | Male | ≥5 years | 5 - 10 years | Low (12) |
| O28 | O-RAG | 50 - 59 | Male | ≥5 years | 5 - 10 years | Minimal (1) |
| C29 | C-RAG | 60 - 64 | Male | 3 - 5 years | 5 - 10 years | Low (4) |
| O30 | O-RAG | 50 - 59 | Male | 3 - 5 years | 5 - 10 years | Low (4) |

[a]O-RAG: open-domain general retrieval-augmented generation.

[b]C-RAG: constraint retrieval-augmented generation.

Although the 5 themes emphasize perceived value, participants also voiced limitations and improvement requests. These included the burden of repeatedly providing clinical context in O-RAG, diminished usefulness for patients receiving long-term dialysis who found responses redundant, and a preference for automatic contextualization. A small number of participants also reported occasional technical issues, such as response delays or repetition of previously provided information, although these were not identified as a dominant theme. We report these alongside the positive accounts to avoid an overly favorable portrayal.

### *Theme 1: Intrinsic Value of Disease-Specific Chatbot Access*

Participants in both groups expressed significant appreciation for immediate access to kidney disease information through the chatbot system, regardless of personalization features. This baseline value was reflected in temporal accessibility, disease domain specificity, and a reduced burden on clinical staff.

> *The answers were helpful, so I continued using it. I could rely on it every time I ate. Its immediate responses—whether I was starting to eat or applying medication—made it very convenient to use.* [O4]

> *It provided more detailed answers than what doctors typically offer, which was both comfortable and beneficial. The ability to ask any questions was very helpful and convenient.* [O1]

Both groups expressed a strong desire to continue using the disease-specific AI assistants after the study, indicating that these tools address fundamental structural limitations in health care delivery.

### *Theme 2: Varied Perception of Personalization Independent of System Design*

Participant perceptions of whether responses were personalized showed substantial variability within both groups and did not consistently align with actual system assignments. Some O-RAG users perceived responses as personalized, despite the generic system design, suggesting that motivated users achieved functional personalization by initiating context in their queries.

> *I recognized it as personalized. When I asked whether I could eat spinach and how much, it responded with the recommended grams to consume and detailed the components, aligning with my information.* [O1]

However, O-RAG users reported that achieving this specificity required providing clinical context with each query, and they noted that automatic contextualization would have alleviated this burden. This automatic contextualization, observed qualitatively, parallels the quantitative finding that C-RAG reduced information needs while O-RAG did not, suggesting that the interaction burden of manual contextualization is one mechanism underlying the divergent change scores.

> *Even without a personalized system, explaining that I was on dialysis before asking allowed me to receive tailored answers. However, it would have been more convenient if the chatbot had already known my situation.* [O1]

One plausible interpretation is that repeated exposure to generic answers, which did not resolve patients' specific concerns, generated dissatisfaction and prompted further questioning—heightening awareness of unmet needs rather than fulfilling them. The high-frequency O-RAG user who attributed elevated usage to iterative follow-up is consistent with this account. Because the O-RAG vs control difference did not reach significance, this remains a hypothesis to be tested in larger samples. One high-frequency O-RAG user (89 questions) explicitly attributed his elevated usage volume to the need for iterative query refinement:

> *Many responses were general, so I needed to ask more specific follow-up questions to obtain answers that matched my situation, which increased the total number of questions.* [O4]

C-RAG participants, in contrast, recognized personalization features when the system referenced their pre-entered clinical data without needing additional prompts:



XSL•FO
RenderX

> *At the beginning of each response, the chatbot displayed my health status first, which made me feel that my condition was accurately reflected in the answers.* [C16]

This automatic contextualization reduced the interaction burden needed to obtain clinically relevant responses.

### *Theme 3: Dietary Management as Predominant Information Need*

Of the nine interview participants, 8 identified dietary management as their most important and frequently addressed information need, regardless of group assignment. This universal emphasis highlights the temporal mismatch between the frequency of dietary decisions and the availability of clinical dietary guidance. This aligns with quantitative results showing large effect size differences between C-RAG and O-RAG in the dietary information need domain (Cohen $d$=−0.86, $P$=.02). This convergence between participants' lived experience and the large observed quantitative effect (Cohen $d$=−0.86) reinforces dietary guidance as both the area of greatest unmet need and the domain of greatest intervention benefit.

> *Dietary management was the hardest to manage. Even now, it's not resolved, and the diet is really difficult. It helped with dietary management, and it was good that it told me exactly what not to eat.* [C3]

> *Dietary management was the most challenging aspect. I had many foods I wanted to enjoy, but I was always worried about portion sizes and the phosphorus and potassium content of those foods. Finding a way to address these concerns proved to be very helpful.* [O1]

### *Theme 4: Dialysis Experience Duration as Dominant Moderator of Utility*

The most notable pattern observed in both intervention arms was the inverse relationship between dialysis duration and perceived chatbot utility. Patients with extensive dialysis experience (>3 y) reported significantly limited benefits, regardless of group assignment, while those with shorter experience found considerable value in both systems.

> *Having been on dialysis for 5 years, I already knew a lot, so my usage wasn't high. Only information I was familiar with was presented, which was disappointing. However, it seems very necessary for those who haven't been on dialysis for long.* [C29, C3]

This pattern suggests that chatbot interventions focus on a specific phase of knowledge acquisition in the chronic disease trajectory, rather than offering continuous value throughout all stages of the disease.

### *Theme 5: Enhanced Confidence Through Verification of Uncertain Practices*

Participants in both intervention groups reported increased confidence in managing their kidney disease through the use of a chatbot. This boost in self-efficacy primarily resulted from their ability to clarify uncertain management practices and resolve ambiguities related to daily self-care decisions.

> *Using the chatbot made it easier for me to practice and boosted my confidence in managing kidney disease. The ability to ask questions about things I hesitated to address allowed me to manage my condition more effectively.* [O1]

> *When I asked about something I was already doing and the chatbot confirmed it was correct, I felt relieved. However, when it advised against a certain action, I reconsidered my behavior.* [C16]

This theme demonstrates that the value of chatbots goes beyond merely providing information; it also encompasses psychological benefits such as reduced uncertainty and enhanced self-efficacy.

## Discussion

### Principal Findings

This pilot RCT successfully met its primary goal of providing preliminary efficacy estimates and effect sizes to guide a fully powered definitive trial of personalized RAG-based chatbots for patients receiving dialysis. In this exploratory pilot, C-RAG showed a greater reduction in information needs than O-RAG in both medication and dietary domains (both Cohen $d$=−0.86, $P$=.02). Given the pilot scale, this finding should be interpreted as a preliminary signal supporting personalization as a candidate design feature, to be confirmed in a fully powered trial. When compared with standard care, C-RAG exhibited medium-to-large effect sizes (medication Cohen $d$=−0.64, dietary Cohen $d$=−0.54). Although these findings were not statistically significant in this small pilot study (n=14 per group; $P$=.13 and $P$=.55, respectively), they exceeded typical benchmarks observed in digital health interventions for chronic disease management (typically Cohen $d$=0.20 - 0.35) [39], indicating potential clinical relevance pending validation in adequately powered trials. In contrast, O-RAG displayed minimal differences from the control group (medication Cohen $d$=0.13, $P$≥.99; dietary Cohen $d$=0.27, $P$=.40; diagnostic Cohen $d$=0.22, $P$≥.99), suggesting that generic RAG systems offer limited additional value over standard care.

Post hoc power analysis based on observed effect sizes suggests that a confirmatory trial with 50 - 55 participants per group would achieve 80% power to detect differences between C-RAG and control at α=.05, assuming similar effect magnitudes and variances. Secondary outcomes, including kidney knowledge, self-efficacy, and adherence, did not show significant differences across groups, which aligns with the short 4-week duration suitable for pilot effect size estimation. These outcomes typically require 12 - 24 weeks to demonstrate detectable changes [40] and will be primary focus areas in a future extended definitive trial.

### Mechanisms: Personalization and Information Need Fulfillment

The most striking finding was the divergent pattern of change between C-RAG and O-RAG participants. C-RAG participants showed reductions in information needs across both medication



XSL•FO
RenderX

(Δ=−0.87) and dietary (Δ=−0.67) domains, indicating improved fulfillment of information needs. In contrast, O-RAG participants showed increases in the same domains (Δ=+0.96 and Δ=+0.56, respectively), while control participants showed minimal change (|Δ|<0.2). This divergence suggests that exposure to general RAG-based information may increase patients' awareness of their unmet information needs without adequately addressing them. However, this trend did not achieve statistical significance in the current pilot sample. In contrast, personalized contextualization seems to bridge this gap by providing responses tailored to individual clinical profiles. Qualitative findings shed light on the mechanisms behind these quantitative patterns. O-RAG users who experienced perceived personalization accomplished this through iterative query refinement, which required more cognitive effort compared to the automatic contextualization offered by C-RAG. One frequent O-RAG user (who asked 89 questions) explicitly noted that his higher usage was due to the need for repeated follow-up questions to obtain answers specific to his situation. For patients with limited health literacy or lower motivation, this additional cognitive load may lead to unresolved information needs and an increased awareness of information gaps—exactly the trend observed in the O-RAG change scores.

The effect sizes (Cohen *d*=−0.86) are significantly higher than those typically seen in digital health interventions [28], underscoring the clinical importance of personalization as a design feature in AI-based patient education systems.

### Dietary Management as the Critical Unmet Need

Evidence from multiple sources consistently indicates that dietary management is the primary unmet information need among patients receiving dialysis [16,24]. In this study, dietary queries accounted for the largest single topic category in system logs, comprising 205 out of 691 (29.3%) entries, while eating-related terms appeared most frequently in keyword analysis with 133 occurrences—more than double the frequency of the next category, medication intake, which had 56 occurrences. At baseline, dietary information needs were rated highly across all 3 groups (mean scores ranged from 3.3, SD 1.1 to 3.9, SD 1.1 on a 5-point scale), reflecting a significant unmet need before any intervention. Qualitatively, 8 out of 9 interview participants identified dietary management as their most important and frequently addressed concern, regardless of group assignment.

Temporal usage patterns further illuminate when dietary information needs arise. A morning peak at 9 AM showed a significant dominance of dietary-related queries, accounting for 36% (19/53) of all inquiries, which aligns with premeal decision-making at the start of the day. Similarly, the evening peak at 9 PM also revealed that dietary queries were the most frequent category, making up 29% (14/48) of the total, indicating that patients reflect on dietary choices after dinner while managing complex nutritional restrictions. Overall, dietary information needs seem to peak around mealtimes, highlighting the importance of on-demand access to dietary guidance outside of scheduled clinical visits. This mismatch between the frequency of dietary decision-making and the periodic availability of clinical dietary guidance represents a key gap that personalized RAG systems are well-suited to address.

These findings support previous research indicating that dietary management poses the greatest self-management challenge for patients with CKD [16,41], and expand the literature by identifying the specific temporal contexts in which these information needs arise.

### Dialysis Experience Duration Drives Engagement, While Demographics Do Not

A consistent pattern observed in both quantitative and qualitative analyses was the inverse relationship between dialysis duration and chatbot utility. Patients undergoing shorter dialysis (1 - 3 y) were more likely to engage with the chatbot, with 90% (9/10) classified as high users, compared with only 27.8% (5/18) of those with longer durations. Qualitative feedback indicated that patients with extensive dialysis experience (>3 y) found limited benefit from the chatbot due to redundant information, while those with shorter durations reported significant value from both systems. This trend suggests that chatbot interventions are particularly beneficial during a critical phase of knowledge acquisition in the chronic disease trajectory—specifically, the period following diagnosis when patients need to quickly learn complex self-management skills. As patients gain disease-specific expertise over time, the additional informational value of a general resource and guidance (RAG) system diminishes.

In contrast, engagement levels were not significantly associated with age (*P*=.89), proficiency with digital devices (mean 3.4, SD 1.0 vs mean 3.3, SD 0.8; *P*=.64), or prior chatbot experience (*P*=.45). This indicates that text-based chatbot systems are broadly accessible to diverse dialysis patient populations, including older adults and those without prior technology experience. This finding supports previous research suggesting that patients with CKD from various demographic backgrounds can effectively engage with digital health interventions [40]. Additionally, the observed sex pattern—where all female participants were classified as high users compared with 41.7% (10/24) of male participants (*P*=.11)—merits further investigation in larger samples, as the consistency of this pattern indicates potential sex differences in information-seeking behavior that could inform targeted implementation strategies.

### Comparison With Prior Work

The current findings contribute to the expanding literature on RAG-based chatbot systems for chronic disease management. Wang et al [27] showed that RAG augmentation improved the quality of LLM responses in diabetes education. Similarly, a recent evaluation of a custom RAG chatbot for type 2 diabetes found that evidence-based responses enhanced health literacy and patient trust. However, neither study compared personalized RAG designs with general ones, nor did they investigate whether exposure to RAG differentially impacted the fulfillment of information needs. This study addresses this gap by isolating personalization as an independent design variable, demonstrating that the method of RAG delivery—not just its presence—affects clinical outcomes. These engagement and trust patterns are consistent with human-centered perspectives on health care AI



XSL•FO
RenderX

adoption, which emphasize that sustained use depends on aligning system behavior with user needs and context rather than on technical capability alone [42].

The different trajectories of information needs observed in this study have not been reported before. Previous RCTs of chatbots, such as the trial on prostate cancer by Baumgärtner et al [28] and the comparison of AI-generated vs expert-generated patient content by Lv et al [39], showed improvements in information outcomes without exploring the conditions under which chatbot exposure might inadvertently increase unmet needs. Our finding that general RAG exhibited a trend toward higher information need scores—although not statistically significant compared with the control group—suggests that the indiscriminate use of AI health information systems may not always be beneficial. While this pattern requires validation in larger samples, it presents a potentially important hypothesis with implications for implementation.

In the context of CKD literature, prior digital interventions have mainly focused on self-management behaviors and quality of life [40], with health literacy identified as a key moderator of outcomes [41]. This study complements existing research by showing that the fulfillment of information needs—a proximal outcome closely linked to health literacy—can be effectively addressed through targeted chatbot interventions, and that personalization is a crucial design element to achieve this goal. Moreover, while previous work on CKD information needs has relied on survey-based assessments [16], this study uses real-time interaction logs and qualitative interview data to provide a more detailed understanding of when and how dietary information needs arise in daily life.

## Limitations

Several limitations should be considered when interpreting these findings. First, the limited sample size and statistical power constrained definitive hypothesis testing. This pilot study enrolled 14 participants per group, which was sufficient for estimating effect sizes but underpowered for testing comparisons between C-RAG and the control. A post hoc power analysis using the observed effect sizes and variance estimates indicated a statistical power of 24% for the C-RAG vs control medication comparison (Cohen $d$=−0.64) and 19% for the dietary comparison (Cohen $d$=−0.54) at $\alpha$=.05. In contrast, the better-powered C-RAG vs O-RAG comparison, with an observed power of 72% for Cohen $d$=−0.86, provides the primary basis for conclusions regarding personalization effects. The observed effect sizes offer robust estimates for planning confirmatory trials (n=50 - 55 per group for 80% power) but cannot establish definitive efficacy. Additionally, the small sample size limits the reliability of subgroup analyses.

Second, the short intervention duration restricted the assessment of downstream outcomes. The 4-week period was sufficient to detect proximal changes in information needs but was inadequate for evaluating downstream effects on knowledge, self-efficacy, and adherence, which typically require 12 - 24 weeks [40]. Whether reductions in information needs translate into lasting clinical benefits—such as improved laboratory values, reduced hospitalizations, and delayed progression—remains to be established through extended trials.

Third, the heterogeneity in usage introduced variance related to adherence. Individual usage ranged from 0 to 89 questions (median 13.0, IQR 4.0 - 40.0), indicating significant variability in adherence that may have diminished group-level effects. This underscores the real-world challenges of implementation and highlights the need for strategies to enhance engagement in future deployments. In addition, the explicit weekly-question encouragement may have influenced usage volume and thus the engagement analyses; observed engagement should therefore be interpreted as usage under active prompting rather than spontaneous adoption.

Fourth, unmeasured moderators may have influenced the outcomes. We did not assess health literacy, which could affect both system usage and benefits [41]. Future studies should measure health literacy and explore adaptive content complexity to accommodate diverse comprehension levels. Additionally, while RAG systems reduce the risk of hallucination compared with baseline LLMs [21], occasional errors may still occur. Future implementations should incorporate uncertainty indicators, source citations, and periodic expert reviews. In addition, the community-sourced comparator responses were authentic forum answers that were not matched to the LLM outputs in length, context, or style, which may have affected the comparison's interpretability.

Fifth, limited generalizability constrains external validity. The study was conducted at 3 tertiary hospitals in South Korea, and the enrolled sample was relatively young and digitally proficient. Findings may therefore not generalize to populations with CKD in other cultural or care settings, or to older, less technologically experienced patients, among whom uptake and benefit could differ. In addition, the Korean-language context and the CKD-specific knowledge base may hinder direct transferability to other languages, cultures, and clinical settings. Adaptation requires not only translation but also cultural customization, integration of regional guidelines, and validation by local experts. Ethical considerations—such as patient autonomy, algorithmic bias, data privacy, and accountability—also demand ongoing attention as AI systems become increasingly prevalent in clinical care. Furthermore, the phase 1 model evaluation was conducted in October 2024. Because commercial LLMs are updated frequently, the specific accuracy, safety, and readability rankings may not reflect current model versions; the contribution of phase 1 lies in the empirical, expert-based selection methodology rather than in fixed model rankings. Relatedly, the Korean-adapted readability indices (KoNLPy-based syllable and word segmentation) provided only relative comparisons across models evaluated on identical Korean text and should not be interpreted as absolute grade-level estimates, given that the underlying Flesch-based formulas were developed for English.

## Future Directions

These findings support 4 high-priority directions for future research and implementation. First, a fully powered confirmatory trial is essential. A multicenter RCT with 50 - 60 participants per group and a follow-up period of 12 - 24 weeks would allow for a definitive evaluation of C-RAG's impact on important outcomes, including disease knowledge, self-efficacy, treatment



XSL•FO
RenderX

adherence, and clinical parameters like serum phosphorus and potassium levels. Extended follow-up would also help determine whether reductions in information needs lead to lasting behavioral changes and improved clinical outcomes. Such trials should incorporate strategies to enhance engagement, such as personalized usage prompts and integration into clinical workflows, to address the usage variability observed in this pilot study.

Second, technical enhancements and broader population applicability are needed. Integrating food recognition technology could facilitate real-time dietary assessments, transforming C-RAG from a mere informational tool into an active behavioral support system. Automated electronic medical record connectivity would alleviate the burden of manual data entry while ensuring that responses accurately reflect the current clinical status. Additionally, adapting content complexity based on user comprehension patterns could further improve accessibility. Expanding C-RAG's use beyond dialysis patients to include patients with earlier-stage CKD (stages 3 - 4) would meet a different set of needs focused on preventing disease progression, where sustained behavioral interventions may have a greater long-term impact [17]. Furthermore, the personalized RAG framework could be applied to other complex chronic conditions that require individualized dietary and medication guidance, such as diabetes and heart failure.

Third, implementation science research is crucial for effective real-world deployment. Successful adoption depends on several factors, including clinician acceptance, workflow integration, reimbursement models, and organizational readiness. Implementation studies that use frameworks like RE-AIM (Reach, Effectiveness, Adoption, Implementation, and Maintenance) can offer actionable guidance for health care systems considering chatbot adoption, particularly regarding reach among older adults and patients with limited digital literacy.

Fourth, evaluating health economics and equity is vital for making informed adoption decisions. Cost-effectiveness analyses should investigate whether enhanced information fulfillment leads to fewer hospitalizations for electrolyte imbalances and reduced routine inquiries between visits. Additionally, assessing equity impacts is essential to ensure that AI-based interventions do not worsen existing disparities in access and outcomes, especially in underserved populations where nephrology expertise is scarce.

## Conclusions

This pilot RCT addressed its primary research question of whether personalized RAG (C-RAG) is more effective in reducing information needs than general RAG (O-RAG) and standard care, while also generating robust effect size estimates for definitive trial design. C-RAG showed a greater reduction in information needs than O-RAG, with a large effect size (Cohen $d$=−0.86) exceeding typical digital health benchmarks. Because this was a pilot trial powered for effect-size estimation rather than confirmatory testing, this result should be regarded as a preliminary estimate. When compared to standard care, C-RAG showed medium-to-large effects (Cohen $d$=−0.54 to −0.64) that, although not statistically significant in this small pilot study, provide promising preliminary estimates and critical parameters for a fully powered confirmatory trial (estimated n=50 - 55 per group).

The differing trajectories of the groups highlight an important insight; general RAG systems tend to increase information needs without meeting them, though this trend needs validation through fully powered trials. In contrast, personalized systems effectively address information gaps by providing automatic clinical contextualization. Dietary management surfaced as the primary unmet need, with dialysis duration—rather than demographics or digital proficiency—being the key predictor of engagement.

These findings lay the groundwork for a definitive multicenter trial with sufficient power, extended follow-up (12 - 24 weeks), and thorough outcome assessments that include knowledge, adherence, and clinical parameters. Essential next steps involve conducting a health economic evaluation, assessing equity impact in underserved populations, and testing cross-disease generalizability to determine whether personalized RAG systems can realize their potential as scalable solutions for chronic disease self-management support.

## Acknowledgments

The authors thank the expert reviewers who provided clinical evaluation of the large language model responses: pharmacists Jo Ga-young, Lee Hyun-jin, Yoon Gyeong-rok, Jeong Jin-kyung, and Jeong In-ae; registered dietitians Lee In-seok, Kim Ji-yoon, Kang Sin-suk, Kim Woo-jung, and Hwang Won-sun; and nephrologists Seung Hyun Han and Ara Ko. Their expert assessments of clinical accuracy and safety formed the basis of the Phase 1 evaluation. We also thank Yeyoung Lee and Mi Young Lee for their assistance with patient recruitment and data collection during the Phase 2 clinical trial at the participating hospital sites.

During the preparation of this manuscript, the authors used Claude (Anthropic) to assist with language editing and formatting. The authors reviewed and edited all content and take full responsibility for the content of the publication.

The MMAS-8 Scale, content, name, and trademarks are protected by US copyright and trademark laws. Permission for use of the scale and its coding is required. A license agreement is available from MMAR, LLC [43].

## Funding

This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through the Agriculture and Food Convergence Technologies Program for Research Manpower Development, funded by



XSL•FO
RenderX

the Ministry of Agriculture, Food and Rural Affairs (MAFRA; grant RS-2024 - 00402136 and grant RS-2021-IP321029). The funders had no involvement in the study design, data collection, analysis, interpretation of results, or the writing of the manuscript.

## Data Availability

The datasets generated and analyzed during this study are not publicly available due to patient privacy and confidentiality restrictions under the institutional review board approval conditions. De-identified data supporting the findings of this study are available from the corresponding author upon reasonable request and with appropriate ethical approval. Requests for data access should include a detailed research proposal and evidence of institutional review board approval from the requesting institution.

## Authors' Contributions

Conceptualization: HJH (lead), KWL (supporting)
Data curation: HJH (lead), JYY (supporting)
Formal analysis: HJH
Funding acquisition: KWL
Investigation: HJH (lead), JYY (equal), JHL (equal), SK (equal), KDY (equal), JPL (equal)
Methodology: HJH (lead), JK (equal), DH (equal)
Project administration: HJH
Resources: JHL (equal), SK (equal), KDY (equal), BS (equal), JPL (equal)
Software: JK (lead), DH (equal)
Supervision: KWL
Validation: HJH
Visualization: HJH (lead), JK (equal), DH (equal)
Writing – original draft: HJH (lead), JK (supporting), DH (supporting)
Writing – review & editing: HJH (lead), JK (supporting), DH (supporting), JYY (supporting), JHL (supporting), SK (supporting), KDY (supporting), JYHP (supporting), BS (supporting), JPL (supporting), KWL (supporting)
KWL and JPL contributed equally as co-corresponding authors.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Phase 1 large language model evaluation workflow. Four large language models (ChatGPT-4o, Gemini 1.5 Flash, Claude 3.5 Sonnet, CLOVA X) were evaluated using 60 patient-generated questions sourced from KidneyCafe, an online kidney disease community. Responses were assessed by 15 clinical experts (5 nephrologists, 5 pharmacists, and 5 clinical dietitians) across three domains: inspection and diagnostics, medications and supplements, and nutrition and diet. Anonymized responses were rated on 5-point Likert scales for clinical accuracy and safety. LLM: large language model; RAG: retrieval-augmented generation.
[PNG File, 440 KB - i-jmr_v15i1e96501_app1.png ]

Multimedia Appendix 2
Semistructured interview guide.
[DOCX File, 16 KB - i-jmr_v15i1e96501_app2.docx ]

Multimedia Appendix 3
Readability scores of large language model responses by domain. Readability was assessed using three metrics: Flesch-Kincaid Grade Level (FKGL), Flesch Reading Ease (FRE), and Gunning Fog Index (GFI). ChatGPT-4o achieved the highest readability across all metrics (FKGL 9.8, FRE 62.4, GFI 11.2), corresponding to a high school reading level, compared to Claude 3.5 Sonnet, Gemini 1.5 Flash, and CLOVA X (FKGL 13.2–14.5, college level). LLM: large language model; FKGL: Flesch-Kincaid Grade Level; FRE: Flesch Reading Ease; GFI: Gunning Fog Index.
[PNG File, 251 KB - i-jmr_v15i1e96501_app3.png ]

Multimedia Appendix 4
Secondary Outcomes at 4 Weeks. Secondary outcomes: group comparisons at 4 weeks for kidney disease knowledge, self-efficacy, and treatment adherence domains.
[DOCX File, 16 KB - i-jmr_v15i1e96501_app4.docx ]

Multimedia Appendix 5



XSL•FO
RenderX

Chatbot usage patterns across temporal dimensions and content domains. (A) Topic distribution of 691 total queries, with dietary management as the predominant category (205/691, 29.3%), followed by diagnosis (22.6%) and medication (17.7%). (B) Top 15 query keywords extracted via morpheme analysis (KoNLPy Okt tokenizer), with eating-related terms and medication intake appearing most frequently. (C) Weekday usage patterns showing 1.38-fold higher engagement on weekdays compared to weekends (13.2 vs 9.6 questions per day), with peak usage from Tuesday to Thursday (14.7–17.1 questions per day). (D) Hourly usage patterns identifying 4 daily peaks at 9 AM, 2 PM, 3 PM, and 9 PM; dietary queries were most frequent at the 9 AM (36%) and 9 PM (29%) peaks. HD: hemodialysis; PD: peritoneal dialysis.
[PNG File, 764 KB - i-jmr_v15i1e96501_app5.png ]

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 1990 KB - i-jmr_v15i1e96501_app6.pdf ]

## References


1. Bello AK, Okpechi IG, Levin A, et al. An update on the global disparities in kidney disease burden and care across world countries and regions. Lancet Glob Health 2024 Mar;12(3):e382-e395. [doi: 10.1016/S2214-109X(23)00570-3] [Medline: 38365413]
2. Park JI, Baek H, Jung HH. Prevalence of chronic kidney disease in Korea: the Korean National Health and Nutritional Examination Survey 2011-2013. J Korean Med Sci 2016 Jun;31(6):915-923. [doi: 10.3346/jkms.2016.31.6.915] [Medline: 27247501]
3. Kalantar-Zadeh K, Jafar TH, Nitsch D, Neuen BL, Perkovic V. Chronic kidney disease. The Lancet 2021 Aug;398(10302):786-802. [doi: 10.1016/S0140-6736(21)00519-5]
4. Mason NA. Polypharmacy and medication-related complications in the chronic kidney disease patient. Curr Opin Nephrol Hypertens 2011 Sep;20(5):492-497. [doi: 10.1097/MNH.0b013e328349c261] [Medline: 21788893]
5. Bonner A, Havas K, Douglas C, Thepha T, Bennett P, Clark R. Self-management programmes in stages 1-4 chronic kidney disease: a literature review. J Ren Care 2014 Sep;40(3):194-204. [doi: 10.1111/jorc.12058] [Medline: 24628848]
6. Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003;27 Suppl 3:S227-S232. [doi: 10.5993/ajhb.27.1.s3.6] [Medline: 14672383]
7. Hu Y. Toward retrieval-grounded evaluation for conversational large language model-based risk assessment. JMIR AI 2026 Mar 12;5:e90759. [doi: 10.2196/90759] [Medline: 41818631]
8. Health online 2013. Pew Research Center. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2026-03-15]
9. Cavanaugh KL. Health literacy in diabetes care: explanation, evidence and equipment. Diabetes Management 2011 Mar;1(2):191-199. [doi: 10.2217/dmt.11.5]
10. Ramakrishnan C, Tan NC, Yoon S, et al. Healthcare professionals' perspectives on facilitators of and barriers to CKD management in primary care: a qualitative study in Singapore clinics. BMC Health Serv Res 2022 Apr 26;22(1):560. [doi: 10.1186/s12913-022-07949-9] [Medline: 35473928]
11. Greer RC, Cooper LA, Crews DC, Powe NR, Boulware LE. Quality of patient-physician discussions about CKD in primary care: a cross-sectional study. Am J Kidney Dis 2011 Apr;57(4):583-591. [doi: 10.1053/j.ajkd.2010.08.027] [Medline: 21131116]
12. Talson MD, Ferreira da Silva P, Finlay J, et al. Patient, caregiver, and provider perspectives on improving provider-patient interactions in hemodialysis: a qualitative study. Can J Kidney Health Dis 2025;12:20543581241309986. [doi: 10.1177/20543581241309986] [Medline: 39758856]
13. Tiu H, Fagerlin A, Roney M, et al. Provider perspectives on chronic kidney disease diagnosis delivery. CN 2018;89(3):205-213. [doi: 10.5414/CN109153]
14. Lederer S, Fischer MJ, Gordon HS, Wadhwa A, Popli S, Gordon EJ. Barriers to effective communication between veterans with chronic kidney disease and their healthcare providers. Clin Kidney J 2015 Dec;8(6):766-771. [doi: 10.1093/ckj/sfv079] [Medline: 26613037]
15. Wright Nunes JA, Wallston KA, Eden SK, Shintani AK, Ikizler TA, Cavanaugh KL. Associations among perceived and objective disease knowledge and satisfaction with physician communication in patients with chronic kidney disease. Kidney Int 2011 Dec;80(12):1344-1351. [doi: 10.1038/ki.2011.240] [Medline: 21832984]
16. Hwang HJ, Kim N, You JY, et al. Harnessing social media data to understand information needs about kidney diseases and emotional experiences with disease management: topic and sentiment analysis. J Med Internet Res 2025 Feb 25;27:e64838. [doi: 10.2196/64838] [Medline: 39998877]
17. Ikizler TA, Burrowes JD, Byham-Gray LD, et al. KDOQI clinical practice guideline for nutrition in CKD: 2020 update. Am J Kidney Dis 2020 Sep;76(3 Suppl 1):S1-S107. [doi: 10.1053/j.ajkd.2020.05.006] [Medline: 32829751]
18. Fraser SDS, Taal MW. Multimorbidity in people with chronic kidney disease: implications for outcomes and treatment. Curr Opin Nephrol Hypertens 2016 Nov;25(6):465-472. [doi: 10.1097/MNH.0000000000000270] [Medline: 27490909]




XSL•FO
RenderX

19. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to Patient questions posted to a public social media forum. JAMA Intern Med 2023 Jun 1;183(6):589-596. [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
20. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nature 2023 Aug;620(7972):172-180. [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
21. Hager P, Jungmann F, Holland R, et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nat Med 2024 Sep;30(9):2613-2622. [doi: 10.1038/s41591-024-03097-1] [Medline: 38965432]
22. Aydin S, Karabacak M, Vlachos V, Margetis K. Navigating the potential and pitfalls of large language models in patient-centered medication guidance and self-decision support. Front Med 2025 Jan 23;12:1527864. [doi: 10.3389/fmed.2025.1527864]
23. Neha F, Bhati D, Shukla DK. Generative AI models (2018–2024): advancements and applications in kidney care. BioMedInformatics 2025 Apr;5(2):18. [doi: 10.3390/biomedinformatics5020018]
24. Padial M, Avesani CM, García-Testal A, Cana-Poyatos A, Lindholm B, Segura-Ortí E. Dietary needs, barriers, and facilitators among patients on hemodialysis and their caregivers: the GoodRENal project in Spain. J Ren Nutr 2025 Mar;35(2):337-343. [doi: 10.1053/j.jrn.2024.08.005] [Medline: 39237029]
25. Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. arXiv. Preprint posted online on May 22, 2020. [doi: 10.48550/arXiv.2005.11401]
26. Ngo NT, Nguyen C, Dernoncourt F, Nguyen TH. Comprehensive and practical evaluation of retrieval-augmented generation systems for medical question answering. arXiv. Preprint posted online on Nov 14, 2024. [doi: 10.48550/arXiv.2411.09213]
27. Wang D, Liang J, Ye J, et al. Enhancement of the performance of large language models in diabetes education through retrieval-augmented generation: comparative study. J Med Internet Res 2024 Nov 8;26(1):e58041. [doi: 10.2196/58041] [Medline: 39046096]
28. Baumgärtner K, Byczkowski M, Schmid T, et al. Effectiveness of the medical chatbot PROSCA to inform patients about prostate cancer: results of a randomized controlled trial. Eur Urol Open Sci 2024 Nov;69:80-88. [doi: 10.1016/j.euros.2024.08.022] [Medline: 39329071]
29. Choi W, Wang S, Lee Y, Oh H, Zheng Z. A systematic review of mobile health technologies to support self-management of concurrent diabetes and hypertension. J Am Med Inform Assoc 2020 Jun 1;27(6):939-945. [doi: 10.1093/jamia/ocaa029] [Medline: 32357368]
30. Mao Y, Lin W, Wen J, Chen G. Impact and efficacy of mobile health intervention in the management of diabetes and hypertension: a systematic review and meta-analysis. BMJ Open Diabetes Res Care 2020 Sep;8(1):e001225. [doi: 10.1136/bmjdrc-2020-001225] [Medline: 32988849]
31. Xie C, Zhu D, Wang Z, Zhang H, Wei Z. Compliance-aware discharge agent for auditable ICU discharge planning: a pilot feasibility study using structured eicu records. 2026. [doi: 10.2139/ssrn.6429758]
32. Zhang T, Kishore V, Wu F, Weinberger KQ, Artzi Y. BERTScore: evaluating text generation with BERT. arXiv. Preprint posted online on Apr 21, 2020. [doi: 10.48550/arXiv.1904.09675]
33. Peng S, He J, Huang J, et al. A chronic kidney disease patient awareness questionnaire: development and validation. PLoS ONE 2019;14(5):e0216391. [doi: 10.1371/journal.pone.0216391] [Medline: 31050683]
34. Lin CC, Wu CC, Anderson RM, et al. The chronic kidney disease self-efficacy (CKD-SE) instrument: development and psychometric evaluation. Nephrol Dial Transplant 2012 Oct;27(10):3828-3834. [doi: 10.1093/ndt/gfr788] [Medline: 22344776]
35. Krousel-Wood M, Islam T, Webber LS, Re RN, Morisky DE, Muntner P. New medication adherence scale versus pharmacy fill rates in seniors with hypertension. Am J Manag Care 2009 Jan;15(1):59-66. [Medline: 19146365]
36. Brooke J. SUS -- a quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry: CRC Press; 1996.
37. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
38. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239. [doi: 10.1136/bmj.i5239] [Medline: 27777223]
39. Lv X, Zhang X, Li Y, Ding X, Lai H, Shi J. Leveraging large language models for improved patient access and self-management: assessor-blinded comparison between expert- and AI-generated content. J Med Internet Res 2024 Apr 25;26:e55847. [doi: 10.2196/55847] [Medline: 38663010]
40. Pal S, Panduragan SL, Said FM, Poddar S. Effect of self-management intervention on Improvement of quality of life in chronic kidney disease patients: a scoping review. TONURSJ 2024 Oct 23;18(1). [doi: 10.2174/0118744346323309241009074442]
41. Billany RE, Thopte A, Adenwalla SF, March DS, Burton JO, Graham-Brown MPM. Associations of health literacy with self-management behaviours and health outcomes in chronic kidney disease: a systematic review. J Nephrol 2023 Jun;36(5):1267-1281. [doi: 10.1007/s40620-022-01537-0] [Medline: 36645651]
42. Gao L, Sherwood J, Aleisa N, Damoah A, Lu Y, Qu X. Human-centered AI agents for healthcare and education: a systematic literature review. . Preprint posted online on Aug, 2025 URL: https://www.researchgate.net/publication/




XSL•FO
RenderX

394312988_Human-Centered_AI_Agents_for_Healthcare_and_Education_A_Systematic_Literature_Review [accessed 2026-09-30]
43. Adherence. Morisky Scale. URL: https://www.moriskyscale.com/ [accessed 2026-09-18]

## Abbreviations

**BERTScore:** bidirectional encoder representations from transformers score
**C-RAG:** constraint retrieval-augmented generation
**CKD:** chronic kidney disease
**FKGL:** Flesch-Kincaid Grade Level
**FRE:** Flesch Reading Ease
**GFI:** Gunning Fog Index
**IRB:** institutional review board
**KDIGO:** Kidney Disease: Improving Global Outcomes
**KDOQI:** Kidney Disease Outcomes Quality Initiative
**KiKS:** Kidney Knowledge Survey
**Kt/V:** dialysis adequacy measure (K: urea clearance, t: time, V: volume of distribution)
**LLM:** large language model
**MMAS-8:** Morisky Medication Adherence Scale-8
**MMLU:** massive multitask language understanding
**O-RAG:** open-domain general retrieval-augmented generation
**RAG:** retrieval-augmented generation
**RCT:** randomized controlled trial
**RE-AIM:** Reach, Effectiveness, Adoption, Implementation, Maintenance
**SUS:** System Usability Scale

*Edited by TDA Cardoso; submitted 07.Apr.2026; peer-reviewed by P Haessner, Y Hu; revised version received 27.Jun.2026; accepted 31.Aug.2026; published 13.Nov.2026.*

*Please cite as:*
*Hwang HJ, Kim J, Heo D, You JY, Lee J, Kwon S, Yoo KD, Yoon Park JH, Suh B, Lee JP, Lee KW*
*Effects of a Personalized Retrieval-Augmented Generation Chatbot on Information Needs in Patients With Chronic Kidney Disease: Mixed Methods Randomized Controlled Trial*
*Interact J Med Res 2026;15:e96501*
*URL: https://www.i-jmr.org/2026/1/e96501*
*doi:10.2196/96501*

© Hee Jeong Hwang, Junghwan Kim, Dongseok Heo, Jeong Yun You, Jeonghwan Lee, Soie Kwon, Kyung Don Yoo, Jung Han Yoon Park, Bongwon Suh, Jung Pyo Lee, Ki Won Lee. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 13.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Original Paper

# Intravenous N-Acetylcysteine in Patients With Dengue Fever With Marked Aminotransferase Elevation: Prospective Comparative Study

Huyen Thi Thu Nguyen[1,2*], MD, PhD; Lien Thi Tran[1,2], MD, PhD; Hieu Thi Nguyen[1], MD; Trang Thi Huyen Phan[1], MD; The Anh Ngo[1*], MD, PhD

[1]Viet Tiep Hospital, Hai Phong, Vietnam

[2]Hai phong University Of Medicine and Pharmacy, Hai Phong, Vietnam

[*]these authors contributed equally

**Corresponding Author:**
Huyen Thi Thu Nguyen, MD, PhD
Viet Tiep Hospital
No 1, Nha Thuong Street, Le Chan Ward
Hai Phong
Vietnam
Phone: 84 912927111
Email: thuhuyenhaiphong@gmail.com

## Abstract

**Background:** Dengue fever can lead to significant hepatic involvement and marked aminotransferase elevation and is associated with increased morbidity and prolonged hospitalization. N-acetylcysteine (NAC) has been proposed as a supportive therapy for dengue-induced liver injury. However, prospective comparative evidence from endemic regions such as Vietnam remains limited.

**Objective:** This study aimed to evaluate the association between intravenous NAC administration and clinical outcomes (primarily length of hospital stay), as well as the secondary effects of NAC on liver transaminase levels (posttreatment aspartate aminotransferase [AST] and alanine aminotransferase [ALT] and their changes from baseline) and platelet recovery in patients with dengue fever with marked aminotransferase elevation (AST or ALT ≥400 U/L).

**Methods:** A prospective, nonrandomized comparative study was conducted at the Department of Tropical Diseases, Viet Tiep Hospital, Hai Phong, Vietnam, from June to December 2024. A total of 128 adult patients with laboratory-confirmed dengue fever and marked aminotransferase elevation (AST and/or ALT ≥400 U/L) were enrolled. Treatment allocation was not randomized and was determined by the treating physicians based on aminotransferase elevation and overall clinical condition. Patients received either intravenous NAC (100 mg per kilogram per day) plus standard supportive care (NAC group; n=68) or standard care alone (non-NAC group; n=60). The primary outcome was length of hospital stay. Secondary outcomes included liver transaminase levels and platelet recovery at discharge.

**Results:** Baseline AST, ALT, bilirubin, and international normalized ratio values were broadly comparable between groups, although several clinical and hematological characteristics differed at baseline. The NAC group had a significantly shorter hospital stay (mean 5.01, SD 1.28 vs 6.12, SD 1.25 days; *P*<.001). The AST and ALT levels were significantly lower in the NAC group (AST: mean 154.83, SD 67.22 vs 184.62, SD 79.88 U/L and *P*=.02; ALT: mean 113.09, SD 33.31 vs 156.35, SD 73.66 U/L and *P*=.004). The posttreatment platelet count was significantly higher in the NAC group (median 31.50, IQR 19.00-50.15 G/L vs 17.00, IQR 8.75-36.23 G/L; *P*=.001). The median increase in platelet count was also greater in the NAC group (15.05, IQR 7.05-26.10 G/L vs 6.95, IQR 4.22-15.43 G/L; *P*=.001). No severe infusion-related adverse events were documented during the study period.

**Conclusions:** In this prospective, nonrandomized comparative study, intravenous NAC administration was associated with shorter hospitalization and lower posttreatment aminotransferase levels in adult patients with dengue fever and marked aminotransferase elevation. No severe infusion-related adverse events were documented during the study period. However, the nonrandomized design and potential for confounding limit causal inference. Larger, multicenter randomized controlled trials are warranted to clarify the therapeutic role of NAC in dengue-related hepatic complications.

***(Interact J Med Res 2026;15:e97990)*** doi:10.2196/97990



XSL•FO
RenderX

**KEYWORDS**

dengue fever; N-acetylcysteine; marked aminotransferase elevation; liver enzymes; prospective study; Vietnam

## Introduction

Dengue fever is a vector-borne viral infectious disease caused by dengue virus (DENV) and transmitted primarily by the *Aedes aegypti* and *Aedes albopictus* mosquitoes [1]. It represents a major global public health challenge, particularly in tropical and subtropical regions.

According to the World Health Organization, dengue cases have increased dramatically in recent decades. In 2023, over 5 million cases and more than 5000 deaths were reported across more than 80 countries and territories. Vietnam alone reported 149,557 cases, including 36 fatalities in that period [2].

Hepatic involvement is one of the most clinically significant complications of dengue fever. DENV-induced liver injury ranges from mild transaminase elevations to fulminant acute liver failure and is associated with higher mortality risk and prolonged hospitalization. Clinical manifestations of hepatic damage include nausea, vomiting, anorexia, abdominal pain, elevated transaminases, and hyperbilirubinemia [3,4].

The pathogenesis of DENV-induced liver damage remains incompletely understood. Proposed mechanisms include direct viral cytopathic effects, hypoxia-mediated injury, and immune-mediated hepatocyte apoptosis [3]. During acute infection, antioxidant enzyme levels, including glutathione peroxidase and glutathione reductase, decline, whereas aspartate aminotransferase (AST) and alanine aminotransferase (ALT) rise [5]. Experimental murine studies have demonstrated that DENV-induced acute liver failure involves proinflammatory cytokines (interleukin [IL]-2, IL-6, IL-8, and IL-22) and that N-acetylcysteine (NAC) attenuates liver damage, reduces hepatocyte apoptosis, and inhibits viral replication in HepG2 cells via interferon-mediated antiviral pathways [6]. NAC also preserved antioxidant enzyme levels and redox balance in DENV-infected mouse livers [7,8].

Supportive care and close clinical monitoring are central to dengue management [9]. NAC is an established antioxidant and glutathione precursor with well-characterized hepatoprotective properties [10,11]. It is recommended by the European Association for the Study of the Liver for acute liver failure from any cause [12] and has been used in dengue-associated acute liver injury in several Asian countries, with consistently positive outcomes reported [13-15]. However, robust randomized controlled trial evidence specifically on DENV-induced hepatic injury is lacking [16]. The 2023 Vietnamese Ministry of Health guidelines recommend NAC for patients with dengue fever with AST or ALT of 1000 U/L or higher [17]. The threshold of AST or ALT of at least 400 U/L was selected for this study to evaluate NAC at an earlier stage of marked aminotransferase elevation. This threshold should be regarded as an exploratory study criterion rather than an established indication for routine NAC treatment. The nonrandomized findings are intended to generate hypotheses and do not support changing clinical guidelines without confirmation in randomized controlled trials.

## Methods

### Study Design and Setting

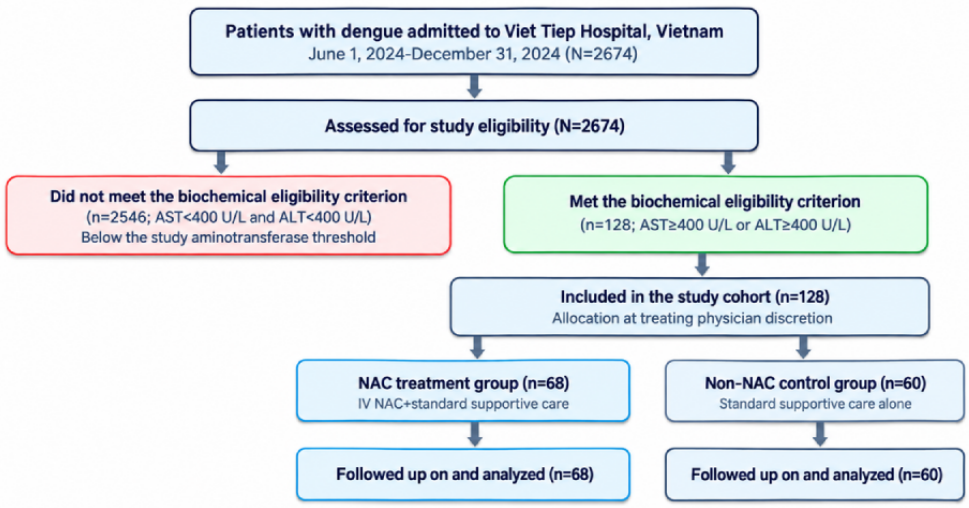
This was a prospective, nonrandomized observational comparative study conducted at the Department of Tropical Diseases, Viet Tiep Hospital, Hai Phong, Vietnam, from June 1, 2024, to December 31, 2024. This study aimed to evaluate the association between NAC administration and clinical outcomes and the safety of intravenous NAC as an adjunctive therapy in adult patients with dengue fever complicated by marked aminotransferase elevation. No a priori sample size calculation was performed. This study used a prospective consecutive convenience sample comprising all eligible adult patients with dengue fever with marked aminotransferase elevation (AST or ALT ≥400 U/L) recruited during the predefined study period from June 2024 to December 2024 (N=128; n=68 in the NAC group and n=60 in the non-NAC group).

The detailed patient flow is shown in Figure 1.



XSL•FO
RenderX

**Figure 1.** Flow diagram of patient screening, inclusion, and analysis. ALT: alanine aminotransferase; AST: aspartate aminotransferase; IV: intravenous; NAC: N-acetylcysteine.




## Participants

A total of 2674 patients with dengue fever were initially screened during the study period (June 1, 2024 to December 31, 2024). Of these, 2546 (95.2%) patients did not meet the biochemical threshold for marked aminotransferase elevation (AST or ALT<400 U/L). Among the 128 (4.8%) patients who met the inclusion criterion, no additional patients were excluded based on predefined exclusion criteria.

Ultimately, 128 patients were allocated to either the NAC group (n=68, 53.1%) or the non-NAC group (n=60, 46.9%) based on physician discretion. All 128 enrolled patients completed follow-up and were included in the final analysis.

Patients were eligible for inclusion if they met all the following criteria: (1) age of 18 years or above; (2) laboratory-confirmed dengue fever, defined as a positive dengue NS1 antigen and/or positive dengue immunoglobulin M antibody; and (3) marked aminotransferase elevation, defined as AST of 400 U/L or higher and/or ALT of 400 U/L or higher (the entry threshold of AST or ALT ≥400 U/L was selected as an exploratory study criterion to evaluate potential benefits of NAC at an earlier stage of marked aminotransferase elevation. This cutoff does not represent an established routine indication for NAC or a definition of acute liver failure, and its exploratory nature is recognized).

Patients were excluded if they met any of the following criteria: preexisting chronic liver disease (chronic hepatitis B or C, alcoholic liver disease, or cirrhosis), history of hepatotoxic drug use prior to admission (excluding standard therapeutic-dose paracetamol for symptomatic fever management), pregnancy or lactation, known hypersensitivity to NAC, or age under 18 years.

## Treatment Protocol

Treatment allocation was not randomized. Eligible patients were allocated to 2 groups based on clinical judgment in routine practice, reflecting real-world clinical decision-making. Intravenous NAC was initiated at the discretion of the treating physician after consideration of aminotransferase elevation and the patient's overall clinical condition [17].

All included patients presented with marked aminotransferase elevation (AST or ALT ≥400 U/L) an exploratory threshold selected for this study. Clinicians tended to initiate intravenous NAC in patients presenting with overt hepatic involvement (eg, hepatomegaly and right upper quadrant tenderness) or rapidly progressive transaminitis accompanied by severe systemic symptoms (eg, persistent vomiting), reflecting a targeted pragmatic approach. Patients with stable transaminitis and milder clinical features were managed through standard supportive care alone [17].

In the NAC group, patients received intravenous NAC (BFS-Depara; Ha Noi CPC1 Pharmaceutical JSC) at 100 mg per kilogram per day diluted in 5% glucose solution and administered via slow intravenous infusion. NAC therapy was initiated at the discretion of the treating physician. Treatment was continued for 3 days, with earlier discontinuation at the discretion of the treating physician based on the patient's clinical and biochemical response.

In the non-NAC (control) group, patients received standard supportive care for dengue fever, including fluid management, antipyretics, and close clinical monitoring, without NAC administration.

Standard supportive care was provided according to the Vietnamese national dengue management guideline and institutional practice [17]. It included clinically guided fluid management; symptomatic treatment where appropriate; serial assessment of vital signs and warning signs; complete blood count monitoring; liver biochemical testing; and clinical surveillance for bleeding, plasma leakage, hemodynamic deterioration, and other dengue-related complications. The supportive treatment was individualized; the intensity and components of supportive care were not identical between groups.



XSL•FO
RenderX

### Clinical and Laboratory Monitoring

#### *Overview*

All patients underwent daily monitoring of clinical symptoms and signs; hematological parameters (complete blood count); and biochemical parameters, including AST, ALT, total bilirubin, and international normalized ratio (when indicated). Liver function tests were recorded at baseline (at NAC initiation or hospital admission for the non-NAC group) and at day 3 of treatment or at hospital discharge, whichever occurred first. Patients were closely observed for adverse reactions to NAC infusion, including hypersensitivity reactions, nausea, vomiting, and anaphylaxis.

#### *Safety Assessment*

Safety outcomes and infusion-related reactions were actively monitored through continuous clinical surveillance during intravenous NAC administration. Monitoring focused on hypersensitivity and anaphylactoid symptoms, including rash, pruritus, respiratory distress, hemodynamic changes, and gastrointestinal intolerance.

### Outcomes

The primary outcome was length of hospital stay (days). Secondary outcomes were changes in serum AST and ALT levels after treatment; platelet count recovery, assessed using the minimum platelet count at the end of the treatment assessment period (day 3 or hospital discharge, whichever occurred first); and change from baseline to this posttreatment measurement. In addition, the change in AST and ALT levels (Δ) was calculated.

### Statistical Analysis

Statistical analysis was performed using the SPSS software (IBM Corp). Continuous variables are expressed as means and SDs or medians and IQRs as appropriate. Categorical variables are presented as frequencies and percentages. Continuous variables were compared using the Student 2-tailed *t* test for normally distributed data or the Mann-Whitney *U* test for nonnormally distributed data. Categorical variables were compared using the chi-square or Fisher exact test.

### Ethical Considerations

The study protocol was approved by the scientific council and ethics committee of Viet Tiep Hospital, Hai Phong, Vietnam. Written informed consent was obtained from all participants or their legal guardians prior to enrollment. All procedures were conducted in accordance with the principles of the Declaration of Helsinki.

## Results

### Patient Characteristics

During the study period, 128 patients with laboratory-confirmed dengue fever and marked aminotransferase elevation (AST and/or ALT ≥400 U/L) met the inclusion criteria. Of these, 68 received intravenous NAC, and 60 received standard supportive care alone.

Baseline demographic characteristics are summarized in Table 1. The mean age was 47.79 (SD 17.43) years in the NAC group and 52.81 (SD 14.11) years in the non-NAC group. Although the mean age was numerically lower in the NAC group, the between-group difference was not statistically significant (*P*=.07). Sex distribution was comparable between groups, with no statistically significant differences.

**Table 1.** Demographic characteristics of study patients (N=128).

| Characteristics | NAC[a] (n=68) | Non-NAC (n=60) | Total |
|---|---|---|---|
| **Sex, n (%)** | | | |
| Male | 25 (36.8) | 28 (46.7) | 53 (41.4) |
| Female | 43 (63.2) | 32 (53.3) | 75 (58.6) |
| Age (y), mean (SD) | 47.79 (17.43) | 52.81 (14.11) | 50.14 (16.10) |

[a]NAC: N-acetylcysteine.

### Baseline Clinical and Laboratory Features

Baseline clinical and paraclinical characteristics are shown in Table 2. All 128 included patients had concurrent elevations in both AST and ALT, and no patient had isolated AST elevation. The NAC group had a longer duration of fever prior to admission (mean 4.57, SD 1.31 vs 3.92, SD 1.41 days; *P*=.007) and higher rates of vomiting (37/68, 54.4% vs 19/60, 31.7%; *P*=.01) and hepatomegaly (37/68, 54.4% vs 12/60, 20%; *P*<.001). The higher prevalence of mucosal bleeding in the non-NAC group represented a baseline imbalance. Its cause could not be determined from the available data, and residual confounding related to bleeding severity may remain. Baseline laboratory parameters, including AST, ALT, total bilirubin, hemoglobin, hematocrit, platelet nadir, white blood cell count, and international normalized ratio, were not significantly different between groups (*P*>.05 for all comparisons), indicating comparable baseline biochemical severity despite more pronounced clinical manifestations in the NAC group.



XSL•FO
RenderX

**Table 2.** Baseline clinical and paraclinical characteristics.

| Characteristics | NAC[a] (n=68) | Non-NAC (n=60) | *P* value |
|---|---|---|---|
| Fever duration (d), mean (SD) | 4.57 (1.31) | 3.92 (1.41) | .007 |
| Right upper quadrant pain, n (%) | 29 (42.6) | 13 (21.7) | .01 |
| Nausea, n (%) | 34 (50) | 25 (41.7) | .35 |
| Vomiting, n (%) | 37 (54.4) | 19 (31.7) | .01 |
| Hemorrhage, n (%) | 28 (41.2) | 35 (58.3) | .053 |
| Mucosal bleeding, n (%) | 17 (25) | 30 (50) | .003 |
| Pleural effusion, n (%) | 33 (48.5) | 21 (35) | .12 |
| Gallbladder wall thickening, n (%) | 36 (52.9) | 24 (40) | .14 |
| Hepatomegaly, n (%) | 37 (54.4) | 12 (20) | <.001 |
| Dengue severity—grade 2, n (%) | 53 (77.9) | 54 (90) | .59 |
| Dengue severity—grade 3, n (%) | 5 (7.4) | 6 (10) | .59 |
| Hemoglobin (g/L), mean (SD) | 163.77 (15.97) | 142.55 (12.05) | .07 |
| Maximum hematocrit (L/L), mean (SD) | 0.44 (0.06) | 0.44 (0.06) | .06 |
| Minimum platelet count (G/L), median (IQR) | 12,40 (5.95-26.17) | 6.00 (2.65-12.38) | .003 |
| White blood cell count (G/L), mean (SD) | 4.95 (3.55) | 3.67 (2.30) | .55 |
| INR[b], mean (SD) | 1.01 (0.08) | 1.02 (0.09) | .27 |
| AST[c] (U/L), mean (SD) | 609.68 (275.10) | 511.53 (159.50) | .09 |
| ALT[d] (U/L), mean (SD) | 425.12 (156.13) | 419.96 (162.93) | .80 |
| Total bilirubin (µmol/L), mean (SD) | 12.45 (4.75) | 13.65 (5.24) | .19 |

[a]NAC: N-acetylcysteine.
[b]INR: international normalized ratio.
[c]AST: aspartate aminotransferase.
[d]ALT: alanine aminotransferase.

## Treatment Outcomes

### *Overview*

Posttreatment outcomes are summarized in Table 3.

**Table 3.** Treatment outcomes at discharge.

| Outcomes | NAC[a] (n=68) | Non-NAC (n=60) | *P* value |
|---|---|---|---|
| Hospital stay (d), mean (SD) | 5.01 (1.28) | 6.12 (1.25) | <.001 |
| Posttreatment AST[b] (U/L), mean (SD) | 154.83 (67.22) | 184.62 (79.88) | .02 |
| Posttreatment ALT[c] (U/L), mean (SD) | 113.09 (33.31) | 156.35 (73.66) | .004 |
| ΔAST (U/L), mean (SD) | 454.85 (261.85) | 326.91 (170.44) | .002 |
| ΔALT (U/L), mean (SD) | 312.03 (154.21) | 263.62 (153.99) | .08 |
| Posttreatment platelet count (G/L), median (IQR) | 31.50 (19.00-50.15) | 17.00 (8.75-36.23) | .001 |
| Change in platelet count (G/L), median (IQR) | 15.05 (7.05-26.10) | 6.95 (4.22-15.43) | .001 |

[a]NAC: N-acetylcysteine.
[b]AST: aspartate aminotransferase.
[c]ALT: alanine aminotransferase.



XSL•FO
RenderX

### *Length of Hospital Stay*

The mean duration of hospitalization was significantly shorter in the NAC group than in the non-NAC group (5.01, SD 1.28 vs 6.12, SD 1.25 days; *P*<.001).

### *Liver Enzyme Recovery*

Mean posttreatment AST levels were significantly lower in the NAC group (154.83, SD 67.22 vs 184.62, SD 79.88 U/L; *P*=.02), as were posttreatment mean ALT levels (113.09, SD 33.31 vs 156.35, SD 73.66 U/L; *P*=.004). The mean change in AST was 454.85 (SD 261.85) U/L in the NAC group vs 326.91 (SD 170.44) U/L in the non-NAC group (*P*=.002). The mean change in ALT was 312.03 (SD 154.21) vs 263.62 (SD 153.99), respectively (*P*=.08).

### *Platelet Recovery*

The posttreatment platelet count was significantly higher in the NAC group (median 31.50, IQR 19.00-50.15 G/L vs 17.00, IQR 8.75-36.23 G/L; *P*=.001). The median increase in platelet count was also greater in the NAC group (15.05, IQR 7.05-26.10 G/L vs 6.95, IQR 4.22-15.43 G/L; *P*=.001). However, because platelet counts were not assessed at uniform predefined time points for all participants, the study could not determine whether NAC accelerated the rate of platelet recovery.

### *Safety*

All patients recovered and were discharged in stable condition. No patients required intensive care unit admission, and no severe infusion-related adverse events were recorded during the study period.

## Discussion

### Principal Findings

In this prospective, nonrandomized comparative study of 128 adult patients with dengue fever and marked aminotransferase elevation, intravenous NAC administration was associated with a shorter duration of hospitalization and lower posttreatment AST and ALT levels compared with standard supportive care alone [17]. The mean hospital stay was approximately 1 day shorter in the NAC group (5.01, SD 1.28 vs 6.12, SD 1.25 days; *P*<.001). Posttreatment AST and ALT levels were also significantly lower in the NAC group. These associations were observed despite the NAC group presenting with more pronounced clinical manifestations at baseline, including longer fever duration and higher rates of vomiting and hepatomegaly. NAC was well tolerated, with no infusion-related adverse events documented.

### Detailed Discussion

The observed association between NAC administration and shorter hospital stay is consistent with prior reports, including the case series by Dissanayake et al [15], that have demonstrated shorter hospitalization among patients with dengue fever with severe hepatitis treated with NAC. A reduction of approximately 1 hospital day may be operationally relevant during dengue outbreaks, particularly in resource-constrained settings. However, the clinical and economic importance of this difference was not directly evaluated and should be confirmed in prospective randomized and health economic studies.

The more pronounced reduction in AST and ALT levels in NAC-treated patients aligns with the known antioxidant and hepatoprotective mechanisms of NAC. DENV infection is associated with oxidative stress, mitochondrial dysfunction, and immune-mediated hepatocyte injury [5,6]. As a glutathione precursor, NAC may mitigate oxidative damage and promote hepatocellular recovery, thereby accelerating the normalization of liver enzymes. Experimental data have also suggested that NAC may exert direct antiviral effects by inhibiting DENV replication via interferon-mediated pathways [6], although viral load was not assessed in the present study.

Alternative explanations for the observed findings must be considered. The NAC group had a longer duration of fever prior to admission (mean 4.57, SD 1.31 vs 3.92, SD 1.41 days; *P*=.007). Patients presenting later in the course of dengue may have been closer to the spontaneous recovery phase; therefore, part of the observed biochemical improvement may reflect the natural disease course rather than NAC administration alone. The higher prevalence of vomiting and hepatomegaly in the NAC group may reflect physician selection of patients with more prominent hepatic manifestations. However, the present study cannot determine whether hepatomegaly modifies the association between NAC and clinical outcomes. This question requires a prespecified subgroup analysis within a randomized trial.

Furthermore, the coexistence of hepatomegaly and marked aminotransferase elevation may represent a more clinically pronounced pattern of dengue-associated hepatic involvement. The potential biological effects of NAC, including restoration of glutathione availability and attenuation of oxidative stress, provide a rationale for further evaluating whether patients with this combined clinical and biochemical phenotype might derive greater benefit from NAC. However, the present observational study was not designed to determine whether hepatomegaly modifies the association between NAC treatment and clinical outcomes. This observation should therefore be regarded as hypothesis generating. Future prospective studies with prespecified subgroup and interaction analyses are warranted to determine whether this phenotype identifies patients who may derive greater benefit from NAC.

Although both groups received care according to the same institutional dengue management framework, the intensity and components of supportive treatment were not quantified in sufficient detail. Differences in fluid management, antiemetic treatment, oral intake, or other supportive interventions may have contributed to the observed outcomes.

Mucosal bleeding was assessed at baseline (upon admission), before treatment initiation. Therefore, the higher prevalence in the non-NAC group (30/60, 50% vs 17/68, 25%) represents a baseline imbalance rather than a treatment effect. As this was a nonrandomized study, the specific reason for this imbalance could not be determined from the available data, and residual confounding related to differences in disease or bleeding severity cannot be excluded.



XSL•FO
RenderX

An interesting finding was the difference in platelet counts between groups. While NAC is not conventionally regarded as a thrombopoietic agent, this effect may be indirect, reflecting reduced systemic inflammation and oxidative stress rather than a direct pharmacological action on platelet production. Hepatic recovery could theoretically influence platelet production through restoration of thrombopoietin synthesis; however, dengue-related thrombocytopenia is multifactorial and may also involve bone marrow suppression, immune-mediated platelet destruction, peripheral consumption, and systemic inflammation. Because thrombopoietin was not measured and platelet counts were not assessed at fully standardized time points, this study cannot establish a mechanistic link between NAC, hepatic recovery, thrombopoietin production, and platelet changes. Given the observational nature of this study, this association warrants cautious interpretation.

NAC was well tolerated in all patients, with no infusion-related adverse events documented, consistent with its established safety profile in paracetamol-induced and non–paracetamol-related acute liver failure [12,18].

The findings may be most relevant to adult patients treated in comparable dengue-endemic hospital settings with access to serial aminotransferase testing and safe intravenous infusion monitoring. However, generalizability is limited by the single-center design, local treatment practices, physician-directed treatment allocation, and selected study population. The results should not yet be interpreted as supporting the routine use of NAC at an aminotransferase threshold of 400 U/L in other health care systems.

To our knowledge, this is among the first prospective comparative studies evaluating intravenous NAC specifically for dengue-associated marked aminotransferase elevation in Hai Phong, Vietnam, contributing to the limited evidence base in this field.

## Limitations

This study has several important limitations.

First, the single-center, nonrandomized design with physician-directed treatment allocation may have introduced selection bias and confounding by indication. No formal randomization sequence or allocation concealment procedure was used, and blinding was not possible.

Second, the sample size was modest, and no a priori sample size calculation was performed; the study used a prospective convenience sample of eligible patients during the predefined study period, which may limit the precision and generalizability of the findings.

Third, several baseline clinical characteristics differed between groups, including fever duration, vomiting, hepatomegaly, and mucosal bleeding. These imbalances, together with differences in supportive treatment and incompletely measured prehospital paracetamol exposure, may have contributed to the observed outcomes, and residual or unmeasured confounding cannot be excluded.

Fourth, supportive treatment was individualized according to clinical condition, and its intensity and specific components were not quantified in sufficient detail. Differences in fluid management, antiemetic treatment, oral intake, or other supportive interventions may therefore have influenced length of hospital stay and biochemical outcomes.

Fifth, platelet counts were not assessed at uniform predefined time points for all participants; consequently, this study could not adequately compare platelet trajectories or determine whether NAC accelerated platelet recovery.

Sixth, secondary biochemical markers, including alkaline phosphatase, gamma-glutamyl transferase, serum albumin, and creatinine, were not systematically measured at uniform time points, limiting a more comprehensive longitudinal assessment of hepatic and renal involvement.

Seventh, viral load and dengue serotype were not measured, precluding assessment of their potential influence on treatment response and clinical outcomes.

Eighth, safety monitoring relied on routine inpatient clinical surveillance rather than a standardized prospective adverse event grading system. Although no severe infusion-related adverse events were documented, mild or transient reactions may have been underrecorded.

Finally, follow-up was limited to the hospitalization period, and longer-term outcomes were not assessed.

## Conclusions

In this prospective, nonrandomized comparative study, intravenous NAC administration was associated with shorter hospitalization and lower posttreatment aminotransferase levels in adult patients with dengue fever and marked aminotransferase elevation. No severe infusion-related adverse events were documented during the study period, although routine clinical surveillance may have underestimated mild or transient reactions. However, the nonrandomized design, potential for physician treatment bias, baseline clinical imbalances, and unmeasured confounding limit causal inference. These findings suggest that NAC may have potential as an adjunctive therapy in this clinical setting. Larger, multicenter randomized controlled trials with standardized supportive care protocols and prespecified end points are warranted to clarify the therapeutic role of NAC in dengue-related hepatic complications and determine whether the observed associations reflect a true treatment effect.

## Acknowledgments

The authors thank the patients, their families, and the clinical teams at Viet Tiep Hospital for their contributions to this study. During the preparation of this work, the authors used an AI-assisted writing tool (ChatGPT 5.5; OpenAI) to check grammar and



XSL•FO
RenderX

improve the readability of the manuscript. No AI tool was used to generate scientific content. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the accuracy and integrity of the publication.

## Funding

This study was funded by the Department of Science and Technology of Hai Phong City. The funder had no role in the study design; data collection, analysis, and interpretation; manuscript preparation; or decision to submit the manuscript for publication.

## Data Availability

The datasets supporting the conclusions of this study are available from the corresponding author (HN) on reasonable request.

## Authors' Contributions

HTTN and TAN designed the study and drafted the manuscript. LTT, HTN, and TTHP oversaw patient recruitment and data collection. TAN and LTT performed the statistical analysis. All authors contributed to interpretation of the data, critically revised the manuscript for important intellectual content, and approved the final version for submission.

## Conflicts of Interest

None declared.

## References

1. Wang WH, Urbina AN, Chang MR, Assavalapsakul W, Lu PL, Chen YH, et al. Dengue hemorrhagic fever - a systemic literature review of current perspectives on pathogenesis, prevention and control. J Microbiol Immunol Infect 2020 Dec;53(6):963-978 [FREE Full text] [doi: 10.1016/j.jmii.2020.03.007] [Medline: 32265181]
2. Disease outbreak news: Dengue – global situation. World Health Organization. 2023 Dec 21. URL: https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON498 [accessed 2026-03-28]
3. Samanta J, Sharma V. Dengue and its effects on liver. World J Clin Cases 2015 Feb 16;3(2):125-131 [FREE Full text] [doi: 10.12998/wjcc.v3.i2.125] [Medline: 25685758]
4. Parkash O, Almas A, Jafri SM, Hamid S, Akhtar J, Alishah H. Severity of acute hepatitis and its outcome in patients with dengue fever in a tertiary care hospital Karachi, Pakistan (South Asia). BMC Gastroenterol 2010 May 07;10:43 [FREE Full text] [doi: 10.1186/1471-230X-10-43] [Medline: 20459677]
5. Chandrasena L, De Silva A, De Mel C, Peiris H, Abesuriya V, De Mel S, et al. Glutathione enzymes and liver injury in acute dengue viral infection. J Biosci Med 2019 Oct;07(10):61-71 [FREE Full text] [doi: 10.4236/jbm.2019.710006]
6. Sreekanth GP, Panaampon J, Suttitheptumrong A, Chuncharunee A, Bootkunha J, Yenchitsomanus PT, et al. Drug repurposing of N-acetyl cysteine as antiviral against dengue virus infection. Antiviral Res 2019 Jun;166:42-55 [FREE Full text] [doi: 10.1016/j.antiviral.2019.03.011] [Medline: 30928439]
7. Guabiraba R, Besnard AG, Marques RE, Maillet I, Fagundes CT, Conceição TM, et al. IL-22 modulates IL-17A production and controls inflammation and tissue damage in experimental dengue infection. Eur J Immunol 2013 Jun;43(6):1529-1544 [FREE Full text] [doi: 10.1002/eji.201243229] [Medline: 23505056]
8. Tafere GG, Wondafrash DZ, Demoz FB. Repurposing of N-acetylcysteine for the treatment of dengue virus-induced acute liver failure. Hepat Med 2020 Nov 3;12:173-178 [FREE Full text] [doi: 10.2147/HMER.S263840] [Medline: 33177895]
9. Dengue guidelines, for diagnosis, treatment, prevention and control. World Health Organization. 2009. URL: https://www.who.int/publications/i/item/9789241547871 [accessed 2026-02-22]
10. Aldini G, Altomare A, Baron G, Vistoli G, Carini M, Borsani L, et al. N-acetylcysteine as an antioxidant and disulphide breaking agent: the reasons why. Free Radic Res 2018 Jul;52(7):751-762 [FREE Full text] [doi: 10.1080/10715762.2018.1468564] [Medline: 29742938]
11. Sahasrabudhe SA, Terluk MR, Kartha RV. N-acetylcysteine pharmacology and applications in rare diseases-repurposing an old antioxidant. Antioxidants (Basel) 2023 Jun 21;12(7):1316 [FREE Full text] [doi: 10.3390/antiox12071316] [Medline: 37507857]
12. European Association for the Study of the Liver. Electronic address: easloffice@easloffice.eu, Clinical practice guidelines panel, Wendon J, Panel members, Cordoba J, Dhawan A, EASL Governing Board representative, et al. EASL clinical practical guidelines on the management of acute (fulminant) liver failure. J Hepatol 2017 May;66(5):1047-1081 [FREE Full text] [doi: 10.1016/j.jhep.2016.12.003] [Medline: 28417882]
13. Dalugama C, Gawarammana IB. Dengue hemorrhagic fever complicated with acute liver failure: a case report. J Med Case Rep 2017 Dec 08;11(1):341 [FREE Full text] [doi: 10.1186/s13256-017-1510-1] [Medline: 29216924]
14. Dalugama C, Gawarammana IB. Lessons learnt from managing a case of dengue hemorrhagic fever complicated with acute liver failure and acute kidney injury: a case report. J Med Case Rep 2018 Aug 08;12(1):215 [FREE Full text] [doi: 10.1186/s13256-018-1766-0] [Medline: 30089506]



XSL•FO
RenderX

15. Dissanayake DM, Gunaratne WM, Kumarihamy KW, Kularatne SA, Kumarasiri PV. Use of intravenous N-acetylcysteine in acute severe hepatitis due to severe dengue infection: a case series. BMC Infect Dis 2021 Sep 20;21(1):978 [FREE Full text] [doi: 10.1186/s12879-021-06681-9] [Medline: 34544380]
16. Gupta M, Gupta S, Sood D, Gupta A, Jesrani G. Role of N-acetylcysteine in liver injury due to dengue fever. Trop Doct 2023 Oct;53(4):475-480 [FREE Full text] [doi: 10.1177/00494755231176317] [Medline: 37312532]
17. Regarding the issuance of guidelines for the diagnosis and treatment of dengue hemorrhagic fever [Report in Vietnamese]. Ministry of Health, Socialist Republic of Vietnam. 2023. URL: https://thuvienphapluat.vn/van-ban/The-thao-Y-te/Quyet-dinh-2760-QD-BYT-2023-Huong-dan-chan-doan-dieu-tri-So-xuat-huyet-Dengue-572227.aspx [accessed 2026-09-25]
18. Chiew AL, Gluud C, Brok J, Buckley NA. Interventions for paracetamol (acetaminophen) overdose. Cochrane Database Syst Rev 2018 Feb 23;2(2):CD003328 [FREE Full text] [doi: 10.1002/14651858.CD003328.pub3] [Medline: 29473717]

## Abbreviations

**ALT:** alanine aminotransferase
**AST:** aspartate aminotransferase
**DENV:** dengue virus
**IL:** interleukin
**NAC:** N-acetylcysteine

*Edited by M Balcarras; submitted 11.Apr.2026; peer-reviewed by S Thapaliya, B Khoundabi, S Arsić; comments to author 09.Jul.2026; revised version received 07.Sep.2026; accepted 15.Sep.2026; published 05.Oct.2026.*

*Please cite as:*
*Nguyen HTT, Tran LT, Nguyen HT, Phan TTH, Ngo TA*
*Intravenous N-Acetylcysteine in Patients With Dengue Fever With Marked Aminotransferase Elevation: Prospective Comparative Study*
*Interact J Med Res 2026;15:e97990*
*URL: https://www.i-jmr.org/2026/1/e97990*
*doi:10.2196/97990*
*PMID:*

©Huyen Thi Thu Nguyen, Lien Thi Tran, Hieu Thi Nguyen, Trang Thi Huyen Phan, The Anh Ngo. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 05.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Retraction

# Retraction: Antibiotic Prescribing Behavior of Physicians in Outpatient Departments in Hospitals in Northwest Ethiopia: Structural Equation Modeling Approach

JMIR Publications Editorial Office[1]

JMIR Publications, Toronto

**Corresponding Author:**
JMIR Publications Editorial Office
JMIR Publications
130 Queens Quay E, Suite 1100
Toronto, Ontario M5A 0P6
Canada
Phone: 1 416 583 2040
Email: ed-support@jmir.org

**Related Article:**

Retraction of: https://www.i-jmr.org/2024/1/e57285

***(Interact J Med Res 2026;15:e105480)*** doi:10.2196/105480

The JMIR Publications Editorial Office is retracting the article "Antibiotic Prescribing Behavior of Physicians in Outpatient Departments in Hospitals in Northwest Ethiopia: Structural Equation Modeling Approach" [1] due to concerns regarding the integrity of the peer review process.

Author Asrat Agalu Abejew responded, stating that there were no conflicts of interest with the reviewers. However, the Editorial Office has determined that there is sufficient evidence indicating that the peer review process was compromised to proceed with a retraction.

Author Asrat Agalu Abejew disagreed with the decision to retract the article. The other authors did not respond to communications.

## Reference

1. Abejew AA, Wubetu GY, Fenta TG. Antibiotic prescribing behavior of physicians in outpatient departments in hospitals in northwest Ethiopia: structural equation modeling approach. Interact J Med Res 2024 Oct 23;13:e57285 [FREE Full text] [doi: 10.2196/57285] [Medline: 39441643]

*Submitted 24.Jun.2026; this is a non–peer-reviewed article;accepted 25.Jun.2026; published 18.Aug.2026.*

*Please cite as:*
*Editorial Office JMIRP*
*Retraction: Antibiotic Prescribing Behavior of Physicians in Outpatient Departments in Hospitals in Northwest Ethiopia: Structural Equation Modeling Approach*
*Interact J Med Res 2026;15:e105480*
*URL: https://www.i-jmr.org/2026/1/e105480*
*doi:10.2196/105480*
*PMID:42612207*

©JMIR Publications Editorial Office. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 18.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete



XSL•FO
RenderX

bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.



XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org



XSL•FO
RenderX